جامعة حلب
كلية الهندسة الكهربائية والإلكترونية
قسم الهندسة الإلكترونية

تصميم نظام صوتي لمساعدة المعاقين في التعامل مع الوسط المحيط

Design of Handi Capped Sound Based System to Control The Environment

أطروحة أعدت لنيل درجة الماجستير في الهندسة الإلكترونية

الدكتور ياسر استانبولي
قسم الهندسة الإلكترونية
كلية الهندسة الكهربائية والإلكترونية
جامعة حلب

الدكتور عمر سردار
قسم الهندسة الإلكترونية
كلية الهندسة الكهربائية والإلكترونية
جامعة حلب

إعداد المهندس رمضان المحمود
طالب دراسات عليا(ماجستر)في قسم الهندسة الإلكترونية
كلية الهندسة الكهربائية والإلكترونية
جامعة حلب

٢٠١٠ م
١٤٣١ هـ

قدمت هذه الرسالة استكمالاً لمتطلبات نيل درجة الماجستير في الهندسة الكهربائية في قسم الهندسة الإلكترونية من كلية الهندسة الكهربائية والإلكترونية بجامعة حلب

أصرح بأن هذا العمل:

"إصدار الأوامر الصوتية من شخص معاق للتحكم بالأجهزة المحيطة"

لم يسبق أن قُبل للحصول على أي شهادة، ولا هو مقدم حالياً للحصول على أي شهادة أخرى.

المرشح

المهندس رمضان المحمود

نوقشت هذه الرسالة بتاريخ ٢٩ / ١٢/٢٠١٠ وأجيزت

لجنة الحكم على الرسالة

الدكتور: لؤي شاشاتي
الأستاذ المساعد في قسم الهندسة الإلكترونية
كلية الهندسة الكهربائية و الإلكترونية
جامعة حلب

الدكتور: محمد فراس حناوي
الأستاذ المساعد في قسم الهندسة الطبية
كلية الهندسة الكهربائية و الميكانيكية
جامعة دمشق

الدكتور: ياسر استانبولي
مدرس في قسم الهندسة الإلكترونية
كلية الهندسة الكهربائية والإلكترونية
جامعة حلب

الفهرس

ملخص

شهدت التسعينيات أول تعامل مع الأنظمة المنطوقة بعد ما بات في وقت من الأوقات حلم مبرمجي الحاسوب الوصول إلى الطريقة التي تمكنهم من التحدث مع الحاسوب والتخاطب معه. هذه الطريقة أصبحت تستخدم لمساعدة من يواجهون صعوبة في استخدام لوحة المفاتيح أو الماوس ليتعاملوا مع الحاسوب، وهي مع ذلك وسيلة للرفاهية والراحة لدى الإنسان. إن الهدف الأساسي من عملية تمييز الصوت تكون في تحديد المعلومات المتواجدة في الإشارة الصوتية وبشكل مستقل عن التغيرات التي تنتج عن هوية الشخص الناطق، وأسلوب الكلام، والشروط المحيطة، وذلك لإيجاد الخوارزمية المناسبة لتوليد البصمة الصوتية لكل كلمة، ومن ثم استخدامها في عملية التحكم.

في هذه الأطروحة قمنا بتصميم نظام لمساعدة المعاقين للتحكم بالأجهزة المحيطة عن طريق الصوت،حيث تم إجراء دراسة للإشارة الصوتية والخوارزميات المتعلقة بعملية تمييز الصوت وتم التوصل إلى الخوارزمية المطلوبة بعد إجراء عملية المحاكاة، وقد طبقت هذه الدراسة من خلال نظام الكتروني.

إذاً لابد من إجراء دراسة شاملة على مفهوم الصوت ومعالجته وهذا ما سنطرق إليه في فصول هذه الرسالة.

الفصل الأول:

بعنوان "أساسيات الإشارة الصوتية" ويتضمن هذا الفصل التعريف بالإشارة الصوتية وكيفية توليد الكلام عند الإنسان وذلك بعرض المخططات الأساسية لتوليد الكلام بالإضافة لآلية انتشار الصوت في جهاز النطق عند الإنسان.

الفصل الثاني:

الفصل الثالث بعنوان " الدراسة النظرية لنظام التعرف على الصوت" في هذا الفصل تم دراسة خوارزميات تمييز الصوت وتصميم الخوارزمية المناسبة لذلك.

الفصل الثالث:

بعنوان " محاكاة باستخدام برنامج Matlab " ، في هذا الفصل تم إجراء محاكاة لخوارزمية تمييز الصوت باستخدام برنامج Matlab.

الفصل الرابع:

بعنوان " الدارة العملية لتمييز الصوت"، في هذا الفصل تم عرض الدارة العملية ونتائجها العملية.

الفصل الخامس:

الخلاصة العامة والآفاق المستقبلية.

مخطط البحث وأهدافه

تأتي أهمية هذا البحث من وجود بعض الأشخاص المعاقين الذين يشكون من إعاقة تامة لكامل أجزاء الجسم والتي تمنعهم من الاستقلالية في اتخاذ قراراتهم في تصرفاتهم ورغباتهم واتصالهم مع الوسط المحيط،هذا بالإضافة إلى أن عمل هذه الأنظمة لا يقتصر فقط على مساعدة المعاقين في الاستقلالية في التعامل مع البيئة المحيطة بهم، بل فتحت مجالاً واسعاً من التطبيقات نذكر منها:

- تفاعل الإنسان مع الحاسب
- المجال الأمني: حيث يمكن استخدام البصمة الصوتية للشخص مثل كلمة مرور في الدخول إلى أي نظام رقمي.

يشمل البحث النقاط التالية:

أولاً: الدراسة النظرية:

- معلومات عامة عن الإشارة الصوتية
- آلية توليد الكلام
- معالجة الصوت
- مبادئ تحليل ومعالجة الإشارة لعملية تمييز الصوت
- الخوارزميات المستخدمة في عملية تمييز الصوت
- تصميم خوارزمية لتمييز الصوت
- إجراء محاكاة باستخدام برنامج Matlab

ثانياً: الدراسة التطبيقية:

إجراء محاكاة للخوارزمية المستخدمة لتمييز الصوت باستخدام برنامج Matlab ، ودراسة النتائج التي تم الحصول عليها، وتنفيذ الدارة الإلكترونية.

مخطط البحث:

- إجراء دراسة للإشارة الصوتية وآلية توليدها وانتشارها.
- دراسة التقنيات المستخدمة في عملية معالجة الإشارة الصوتية.
- دراسة الخوارزميات المستخدمة في عملية تمييز الصوت.
- تصميم الخوارزمية التي تم استخدامها في عملية تمييز الصوت.

أهمية البحث وأهدافه:

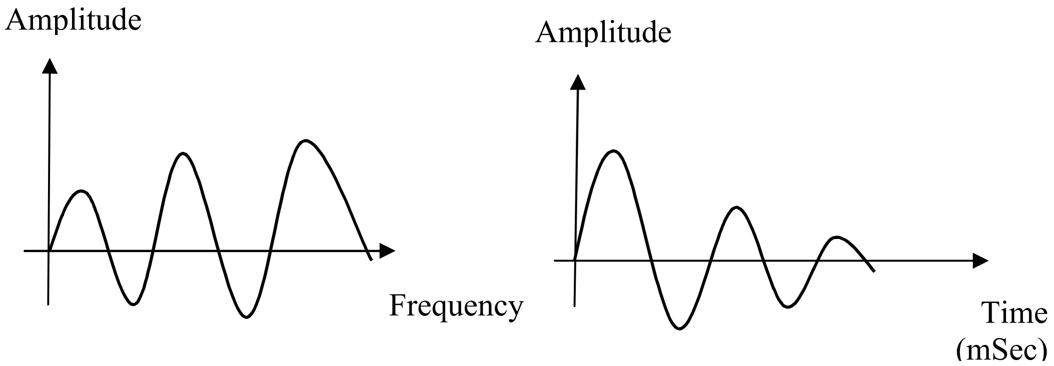
يهدف هذا البحث إلى تصميم وتطبيق نظام تمييز صوتي بهدف مساعدة الإنسان المعاق للتحكم بالبيئة المحيطة،بحيث يعطي الشخص المعاق الأوامر الصوتية إلى النظام ،الذي بدوره يقوم بتحليل هذه الأوامر ويقوم بالتحكم بإحدى الأجهزة المحيطة مثل كرسي المعاقين الذي يسير على العجلات،التلفاز، الأبواب ...الخ.

الفصل الأول

أساسيات الإشارة الصوتية

١-١ معلومات عامة عن الإشارة الصوتية

في عمليات التعرف على الكلام من المهم أن نركز على كيفية إنتاج الكلام والتي تبدأ بالتفكير في الرسالة التي يريد الشخص النطق بها وهذه الرسالة ستتحول بعد ذلك إلى إشارات عصبية والتي ستصبح غالبا إشارات كهربائية تحرك أعضاء النطق، بعد أن تتحرك أعضاء النطق سوف تنفذ سلسلة متتابعة من الإشارات التي ينتج في نهايتها الموجة الصوتية التي تحتوي على معلومات الرسالة الأصلية وعندها تسمع الرسالة للطرف الآخر ..أي سوف تنتج الكلمة موجة صوتية. .والتي هي عبارة عن شكل من أشكال الموجات وإحدى الطرق التي نستطيع تمثيل الكلام بواسطتها وكل الموجات الصوتية هي عبارة عن إشارة والشكل (١-١) يبين موجة صوتية.



الشكل (١-١) تمثيل الإشارة

وإحداثيات الموجة الصوتية هي المحور الأفقي ويمثل الزمن (Time) المحور الشاقولي ويمثل المطال (Amplitude) وهناك أنواع أخرى للموجات نستطيع بواسطتها تمثيل الكلام ومنها ما يسمى بطيف الإشارة وإحداثياته كالتالي: الأفقي يمثل الترددات (Frequencies) والشاقولي يمثل المطال (Amplitude) وهي أسهل في عملية المعالجة. أما في التحليل الطيفي فإننا نستطيع تمثيل الكلام بها وهي ذات ثلاث أبعاد: الزمن، التردد، المطال ولكنها صعبة في التعامل مع الموجة ومعالجتها [1].

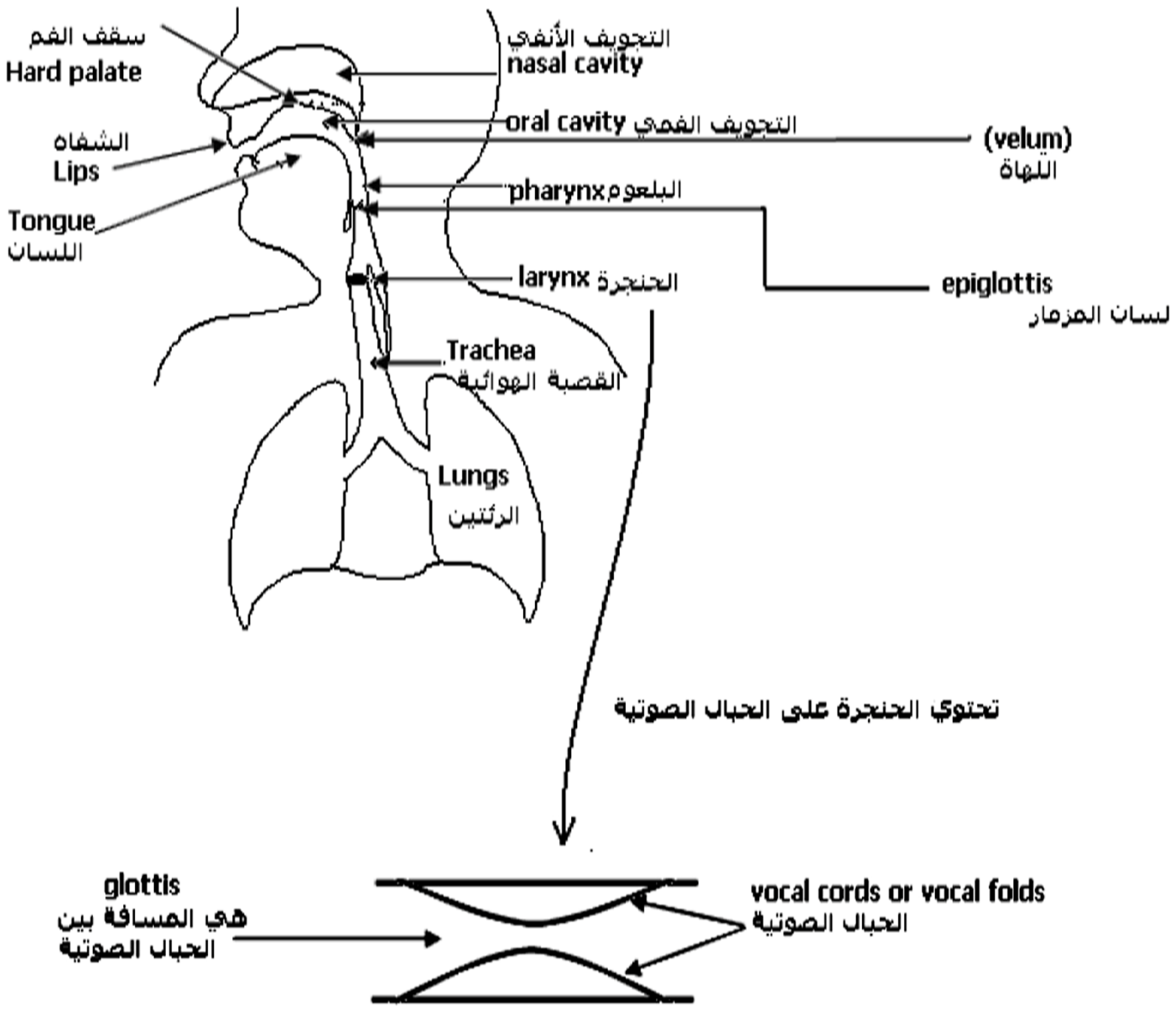
١-٢ آلية توليد الكلام

عندما نريد النطق بأي كلمة في البداية سيتم التنفس من الرئتين ثم يخرج الهواء إلى الحنجرة والذي يسبب عند مروره بها اهتزاز الحبال الصوتية ويستمر في مسيره حتى يصل إلى القناة الصوتية ومنه إلى التجويف الفموي و في بعض الحروف مثل م – ن يمكن أن يضطر الهواء للمرور بالتجويف الأنفي وبعدها تخرج الكلمة وتصبح مسموعة وتتحول إلى موجة صوتية.

إذاً نطلق على الكلمة عندما تخرج من الفم وتتحول إلى موجة صوتية مسموعة مصطلح الكلمة الصوتية.

والعملية العكسية هي الأذن التي تستقبل هذا الكلام حيث تدخل الإشارة إلى طبلة الأذن وتبدأ المطرقة بالضرب وتتصل الرسالة إلى الشخص الآخر ويسمع الكلمة[1] .

الجهاز الصوتي أو جهاز النطق: يتمثل في ثلاث أعضاء رئيسية هي: الحلق والفم والأنف.
الحلق يحتوي على الحبال الصوتية التي تهتز لتنتج لنا الأصوات. وفي الشكل (١-٢) تفصيل لجهاز النطق:

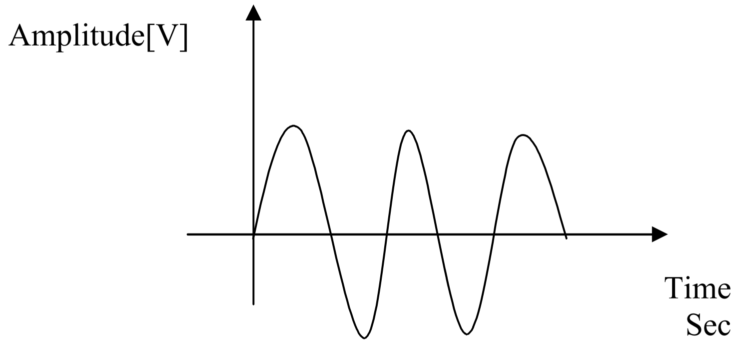


الشكل (١-٢) جهاز النطق عند الإنسان

نقاط النطق تتمثل بـ: الأسنان واللسان واللهاة والجزء العظمي خلف الأسنان العليا والحنك الأعلى أو ما يعرف بـ سقف الفم والشفاه [2].

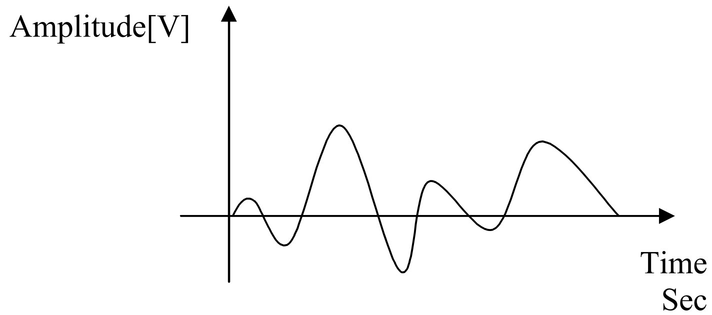
نود أن نوضح الآن عدة مفاهيم تختص بالموجات الصوتية:

عندما تكون الموجات كالشكل(١-٣) حيث الموجات تتكرر بشكل دوري جدا وهو ما نسميه بموجة دورية وهذه الموجات مستحيل أن تعبر عن الموجات الصوتية حيث أن الموجات الصوتية المسموعة موجات غير متكررة [1].



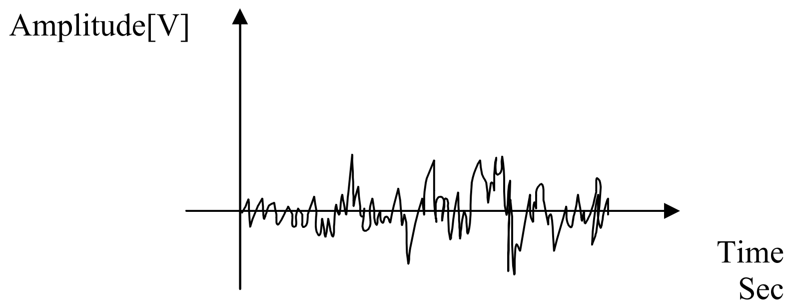
الشكل(١-٣) يمثل الموجة الدورية

ولذلك فإن موجات الصوت المسموعة نمثلها بما يسمى بشبه دورية وهي تعني أنها ليست دورية متساوية مثل الشكل السابق وذلك لكون موجات الكلام عبارة عن مجموعة من الترددات بمطالات مختلفة، ويتخللها فترات من الضجيج وشكلها تقريبا كما في الشكل (١-٤) [1].



الشكل (١-٤) الموجة غير الدورية

وهناك موجات يمثل به الضجيج وذلك بسبب عشوائيتها وهي التي تمثل فترات بين الكلام مثل التنفس كما في الشكل (١-٥)[1] .

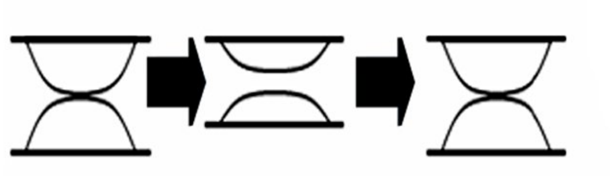


الشكل (١-٥) الضجيج العشوائي

وظيفة الحبال الصوتية تكون عند إخراج الكلام: نستطيع تقسيم حالات الحبال الصوتية إلى ثلاث حالات مختلفة بعد أن نتفق على أن الحبال الصوتية في الأحوال العادية تكون مقفلة نوعا ما.

١- في الأصوات المسموعة (الحبال الصوتية مقفلة) عندها سوف تجبر الحبال الصوتية بواسطة ضغط الهواء الآتي لها من الرئتين والمسبب لإرتطامها ببعض، أن تقوم بعمل شيء من الاهتزاز ثم تعود مرة أخرى إلى حالتها المقفلة تتكرر هذه الحركة أثناء إنتاج الكلام بشكل دوري مما يسبب إنتاج طنين معين

الذي بدوره يسبب إنتاج التردد وطبعا طالما الهواء لم يتعدى الحبال الصوتية لن يتم إنتاج الصوت حتى يتعداها كما في الشكل (١-٦) [1].



الشكل (١-٦) يمثل حالة الحبال الصوتية

وقيمة التردد التي نتجت تعتمد على عدة عوامل من أهمها:

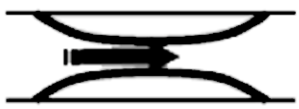
- طول الحبال الصوتية: حيث أنه كلما قصرت الحبال الصوتية كان التردد فيها أعلى.
- كتلتها: حيث أنه إذا كانت كتلتها كبيرة يكون التردد أقل والعكس صحيح.
- مقدار الشد فيها: حيث يزداد التردد كلما زاد الشد في حبال الصوت.

والتردد الناتج عن اهتزاز الحبال الصوتية عادة يتراوح بين 60-400 هرتز حيث يبلغ متوسط التردد في الحبال الصوتية عند النساء البالغات 180Hz وعند الرجال البالغين 100Hz.

٢- الحالة الثانية للحبال الصوتية هي ما تسمى بالصوت الضعيف وليس المعنى أنها ليست مسموعة تماما ولكنها حروف تنطق عند تمثيلها نلاحظ أنها تشبه حالة الضجيج أي ذات قيمة منخفضة عندها ستكون الحبال الصوتية مفتوحة قليلا عن بعضها البعض.

٣- الحالة الثالثة هي كون الحبال الصوتية مفتوحة بشكل كبير أي تكون المسافة بينها واسعة وهذه تحدث أثناء التنفس حيث يسبب انفتاحها الواسع مرور الهواء بينها دون عمل اهتزاز على الحبال الصوتية ولا ينتج صوت حينها.

عندما يكون الشخص لديه التهاب في الحبال الصوتية فإن ذلك يؤدي إلى زيادة سمكها وبالتالي تقل المسافة بين الحبال الصوتية مما يضيق مكان مرور الهواء فيها وبالتالي عندما يتنفس الشخص ينتج صوت ملحوظ للنفس وذلك بعد أن ضاقت المسافة بين الحبال الصوتية فإن الهواء أثناء مروره يسبب اهتزاز للحبال الصوتية مما أدى إلى ظهور الصوت المزعج أثناء التنفس كما في الشكل (١-٧) [1].



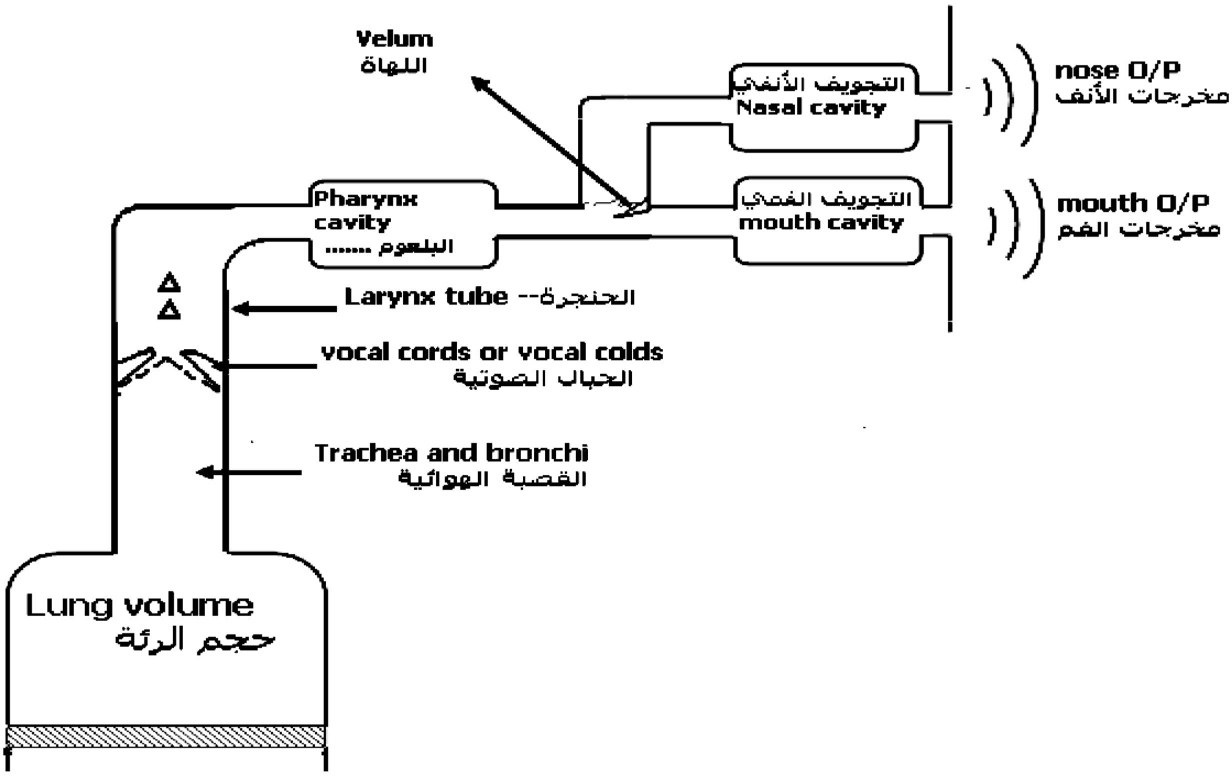
الشكل (١-٧) يمثل حالة التهاب الحبال الصوتية

بعض التعاريف الهامة:

الوحدة الصوتية: وهي عبارة عن مجموعة الرموز التي يتكون منها الصوت ويمكن أن تكون حرف وقد تكون أقل منها.

علم الأصوات: وهو الذي يدرس الصوتيات وتصنيفها ...مثلا تصنف الكلمات المسموعة من الكلمات غير المسموعة وغيرها.

لننتقل الآن إلى آلية إنتاج الكلام وفق المخطط العام لتوليد الصوت ويمكن توضيحه في الشكل (١–٨) [2].



الشكل(١–٨) المخطط العام لتوليد الصوت

نلاحظ أن حجم الرئة يتناسب مع حجم الهواء الذي يكون بداخلها.

نستطع تقسيم الأصوات التي تستطيع القناة الصوتية إخراجها إلى الأحرف الساكنة والأحرف الصوتية.

الأحرف الصوتية: وهي الحروف التي يتم إنتاجها عندما يكون الفم مفتوح قليلا وهي الحروف المدية وهنا تهتز الحبال الصوتية كليا لإنتاجها مثل الـ a, e, u, o, aa, ee, au,uh.

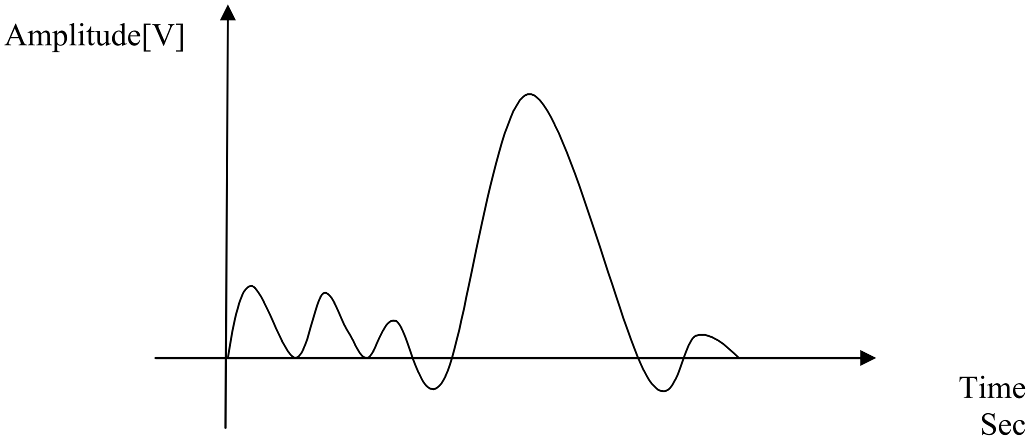
أما الأحرف الساكنة:. وهي ما نسميها بالسواكن وهنا الحبال الصوتية إما أنها لا تهتز كليا أو أنها تهتز بشكل جزئي وطبعا هذه السواكن تعتمد على الحرف الذي يسبقها والحرف الذي يليها[3] .

وهذه الأحرف الساكنة يتم توصيفها اعتمادا على نقطتين:

١– مخرج الحرف:التي قد تكون من الشفاه أو اللسان أو طرف اللسان أو باشتراك بعض نقاط النطق لإخراج هذه الحروف.

٢– طريقة نطق الحرف:والتي تتدرج تحتها النقاط التالية:

- الأحرف الانفجارية: وهي الحروف الانفجارية حيث يتم قفل نقطة من نقاط النطق نهائيا مثل الشفاه أو غيرها ثم بعدها يخرج الحرف فجأة مثل حرف B في كلمة(Bad) وحرف الـ G في كلمة(Gat) وحرف الألف في كلمة (قال) وتمثل الموجة بالشكل (١-٩)



الشكل (١-٩) يمثل الحروف الانفجارية

- الحروف الاحتكاكية: وهي الناتجة عن احتكاك نقطتين من نقاط النطق أثناء النطق بحرف معين مثل الحروف التالية: v, th, z, f, s, s
- الحروف الأنفية :وهي مثل m, n, ng والتي يشارك الكهف الأنفي بجزء كبير من إخراجها.
- الحروف المقفلة: وهي الحروف التي تخرج مع إقفال أحد نقاط النطق تماما ثم يتدرج الهواء في الخروج بعد ذلك مثل حرف J في كلمة (Judge)
- الحروف التقاربية: التي تعني اقتراب أعضاء النطق من بعضها عند إنتاج الحرف ولها نوعين
  - **الحروف المدية:** وهي مثلW, Y
  - **الحروف المنسابة:** وهي مثل R, L [3].

١-٣ انتشار الصوت

في النظرية الصوتية المفصلة يجب أن نأخذ بعين الاعتبار التأثيرات التالية:

١- زمن تغير شكل المسار الصوتي.

٢- نعومة جدران المسار الصوتي.

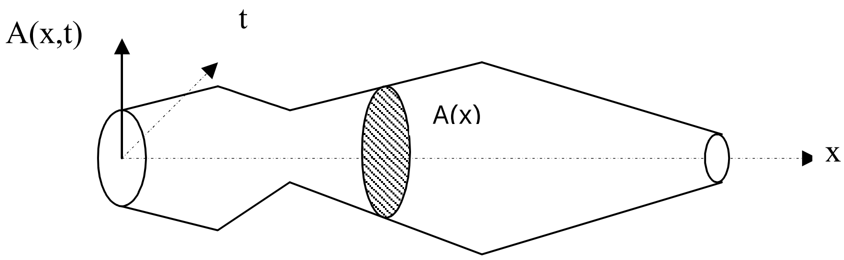
٣- إشعاع الصوت عند الشفاه.

٤- تهييج الصوت في المسار الصوتي.

٥- الفتحات الأنفية.

٦- الاحتكاك اللزج عند جدران المسار الصوتي.

لنبدأ في حالة بسيطة لأنبوب ذو خسارة قليلة كما في الشكل (١-١٠).



الشكل (١-١٠) انتشار الصوت ضمن انبوب

بالنسبة للترددات الصوتية التي طول موجتها اقل من أبعاد المنطقة الصوتية (الترددات الأقل من٤٠٠٠ هرتز) يمكن تطبيق المعادلات التالية:

$$\left.\begin{aligned} &\rho \frac{d(u/A)}{dt} + grad(p) = 0 \\ &\frac{1}{\rho C^2}\frac{dp}{dt} + \frac{dA}{dt} + div(u) = 0 \end{aligned}\right\} \qquad (١-١)$$

$$-\frac{dp}{dx} = \rho \frac{d(u/A)}{dt} \qquad (١-٢)$$

$$-\frac{du}{dx} = \frac{1}{\rho C^2}\frac{d(pA)}{dt} + \frac{dA}{dt} \qquad (١-٣)$$

حيث P: يمثل تغير الضغط في الأنبوب.

u: التغير في سرعة الصوت.

ρ: كثافة الهواء في الأنبوب.

C: سرعة الصوت (35000 cm/s).

A: يمثل تابع المسافة.

وفي حالة A(x,t)=A عندها يمكن أن نكتب المعادلات التالية:

$$-\frac{dp}{dx} = \frac{\rho}{A}\frac{dp}{du} \qquad (١-٤)$$

$$-\frac{du}{dx}=\frac{A}{\rho C^2}\frac{dp}{dt}$$ (١-٥)

يمكن مقارنة هذه القيم الصوتية مع القيم الكهربائية كما يلي:

$$\left.\begin{aligned}-\frac{di}{dx}&=c\frac{dv}{dt}\\ -\frac{dv}{dx}&=L\frac{di}{dt}\end{aligned}\right\}$$ (١-٦)

مقدار صوتي	مقدار كهربائي
مقدار صوتي	**مقدار كهربائي**
ضغط P	الجهد V
سرعة الصوت U	التيار I
المقاومة الصوتية ρ / A	الممانعة L
السعة الصوتية $A/\rho c^2$	السعة الكهربائية C

بحل المعادلتين التفاضليتين (١-٤) و (١-٥) نحصل على ما يلي:

$$p(x,t)=jZ_0\frac{Sin[\Omega(l-x)/c]}{Cos[\Omega l/c]}U_G(\Omega)e^{j\Omega t}$$ (١-٧)

$$U(x,t)=\frac{Cos[\Omega(l-x)/c]}{Cos[\Omega l/c]}U_G(\Omega)e^{j\Omega t}$$ (١-٨)

حيث Ω تمثل التردد الزاوي سوف يتم حسابها.

$$Z_0=\frac{\rho c}{A}$$ (١-٩)

حيث Zo تمثل الممانعة المميزة.

l: طول المسار الصوتي.

ويعطى تابع الانتقال للأنبوب بالمعادلة التالية:

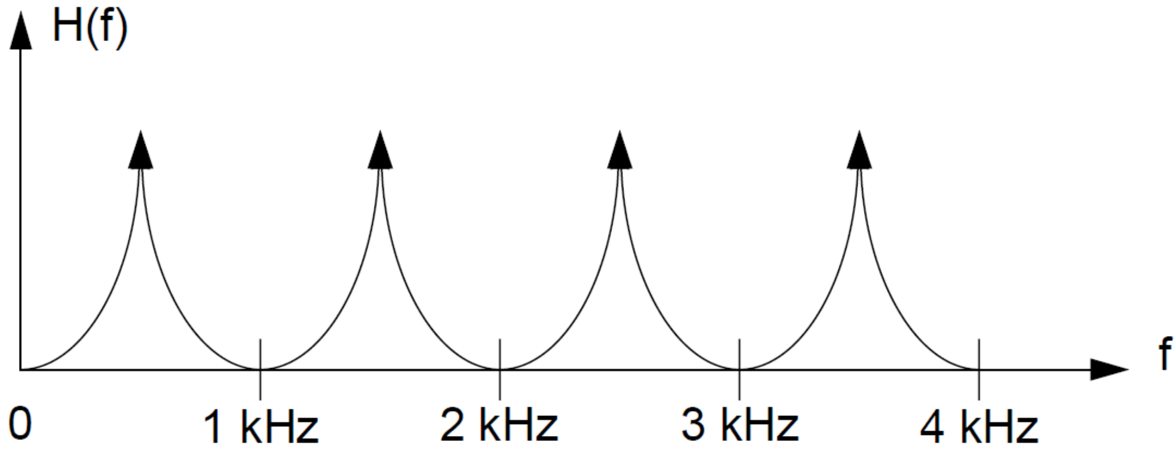
$$\frac{U(l,\Omega)}{U(0,\Omega)}=\frac{1}{Cos(\Omega l/c)}$$ (١-١٠)

هذا التابع يحتوي على أقطاب كل الدور كالتالي:

$$\frac{(2n+1)\pi c}{2l} \qquad (١١-١)$$

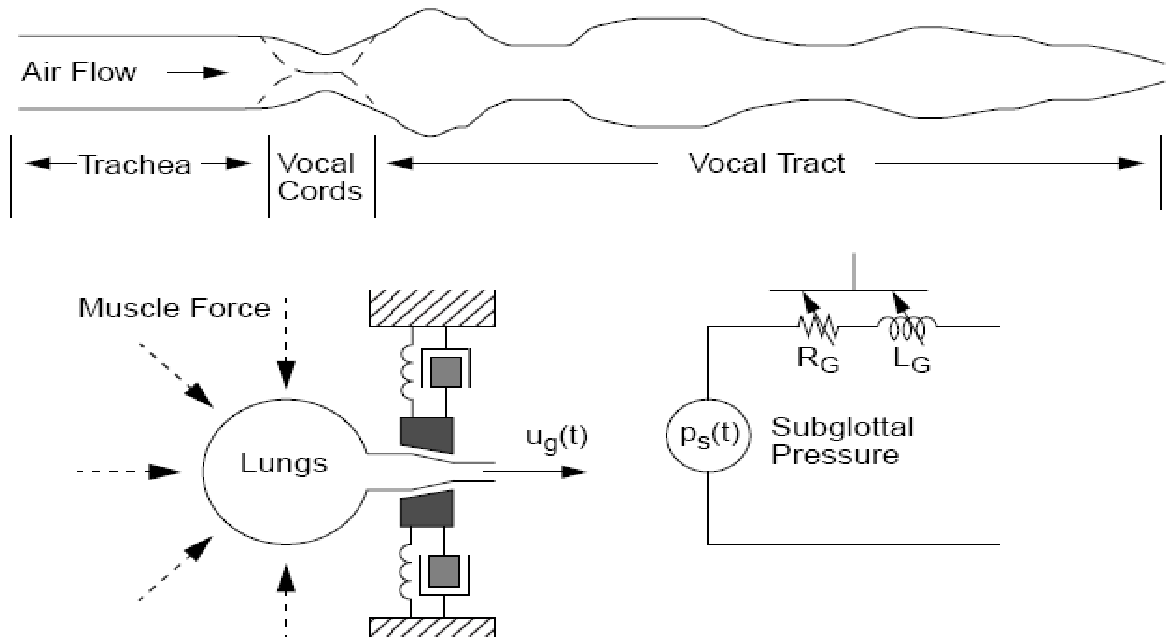
وهذا يتطابق مع الترددات التي يصبح فيها طول الأنبوب مساوياً لربع طول الموجة كما في الشكل (١-١١)

$$\left(\frac{\Omega l}{c}=\frac{\pi}{2}\right)\Rightarrow\left(\Omega=\frac{\pi c}{2l}\right) \qquad (١٢-١)$$



الشكل (١-١١) توافقيات انتشار الصوت ضمن الأنبوب

وبالتالي يمكن الحصول على النموذج الرياضي للمسار الصوتي كما في الشكل (١-١٢)



الشكل (١-١٢) يمثل النموذج الرياضي لانتشار الصوت

ممانعة الحلق تعطى بالعلاقة التالية:

$$Z_G = R_G + j\Omega L_G \qquad (١-١٣)$$

ويمكن الحصول على سرعة الصوت بدلالة الضغط عند الشروط الحدية كما يلي :

X=0,t=Ω

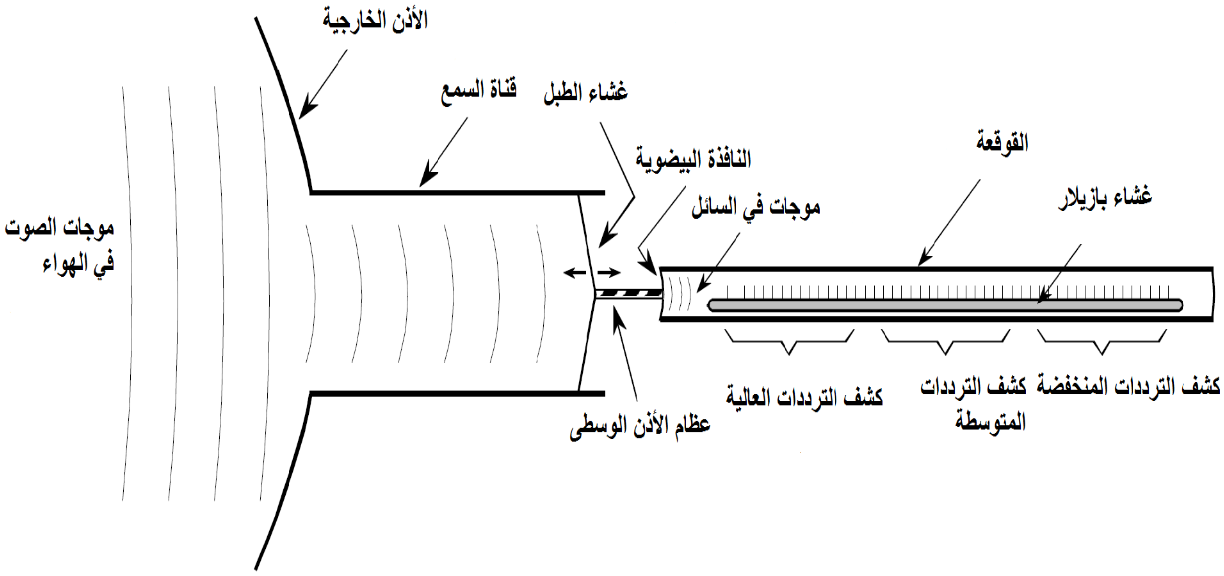
$$U(0,\Omega) = U_G(\Omega) - P(0,\Omega) / Z_G(\Omega) \qquad (١-١٤)$$

[9]

١-٤ معالجة الصوت

١-٤-١ السمع عند الإنسان

تعتبر أذن الإنسان من الأعضاء الشديدة التعقيد و مما يزيد المسائل صعوبة أن المعلومات من كلتا الأدنيين يتم ضمها في شبكة عصبية معقدة وهي الدماغ. نذكر بأن دراستنا التالية هي دراسة عامة و موجزة فهناك العديد من التأثيرات الدقيقة و الظواهر المتعلقة بالسمع عند الإنسان مازالت غامضة يوضح الشكل (١-١٣)البنية الرئيسية لأذن الإنسان و العمليات التي تشكل حاسة السمع عنده، كما هو ملاحظ تتكون الأذن الخارجية (الجزء الظاهر من الأذن) من جزئيين : صيوان الأذن مع غضروف يربطه بالرأس و قناة السمع وهي عبارة عن أنبوب قطره حوالي 0.5 Cm و طوله 3 Cm يمتد داخل الرأس. تقوم الأذن الخارجية بتوجيه الأصوات من الوسط المحيط إلى الأذن الوسطى الحساسة و الأذن الداخلية المتوضعتين بأمان داخل تجويف عظمي في نهاية قناة السمع حيث يوجد غشاء من نسيج رقيق يسمى غشاء الطبل أو طبلة الأذن يهتز هذا الغشاء تحت تأثير الاصطدام بالأمواج الصوتية. أما الأذن الوسطى فهي تركيبة من العظام تنقل الاهتزازات إلى الأذن الداخلية (قوقعة الأذن) التي تحولها إلى نبضات عصبية والأذن الداخلية عبارة عن أنبوب مملوء بسائل يبلغ قطره 2 mm بينما طوله حوالي3Cm يظهر هذا الأنبوب في الشكل على أنه مستقيم إلا أنه في الحقيقة يلتف للأعلى و يكون شكله شبيها بقوقعة حلزونية .



الشكل (١-١٣) جهاز السمع عند الإنسان

عندما تحاول الموجة الصوتية المرور من الهواء إلى السائل سيقوم السطح الفاصل بتمرير جزء صغير فقط من الصوت بينما الجزء الباقي من الطاقة يرتد منعكسا و ذلك لأن الهواء ذو ممانعة ميكانيكية ضعيفة بينما تكون الممانعة الميكانيكية للسائل عالية. يمكن تشبيه هذا بمجالات أقل تقنية فمثلا عند تحريك يدك في الماء ستحتاج إلى طاقة أكبر مما تحتاجه لتحريكها في الهواء سيؤدي هذا الاختلاف بالممانعة الميكانيكية إلى انعكاس الجزء الأكبر من الصوت عند السطح الفاصل بين الهواء و السائل. تقوم الأذن الوسطى بدور شبكة موافقة الممانعة

فتعمل على زيادة الجزء (الاهتزازات) الذي يعبر إلى السائل الذي يملأ الأذن الداخلية و هذا يفسر عدم امتلاك السمك غشاء طبل أو أذن وسطى فهو ليس بحاجة للسمع في الهواء أما عملية تحويل الممانعة فتأتي من اختلاف المساحة بين طبلة الأذن (التي تتلقى الصوت من الهواء) و النافذة البيضوية (التي تنقل الصوت إلى السائل) حيث تبلغ مساحة طبلة الأذن حوالي $60\ mm^2$ بينما مساحة النافذة البيضوية هي $4mm^2$ وبما أن الضغط يساوي ناتج قسمة القوة على المساحة فإن هذا الاختلاف في المساحة سيزيد ضغط الموجة الصوتية بمقدار 15 مرة. يشكل غشاء البازيلار (basilar membrane) الموجود داخل قوقعة الأذن بنية ربط لحوالي 2000 خلية حساسة تكون أعصاب القوقعة. من صفات هذا الغشاء أنه مرن بالقرب من النافذة البيضوية ويصبح أكثر ثباتا باتجاه النهاية الأخرى مما يسمح له بأن يعمل كمحلل طيف عندما يتعرض لاهتزازات عالية التردد فإنه سيهتز عند الجزء المرن مما يؤدي إلى تهييج خلايا العصب القريبة من النافذة البيضوية بنفس الشكل ستهيج الأصوات المنخفضة التردد خلايا العصب الواقعة عند النهاية البعيدة للغشاء. إن هذه الآلية ستجعل شعيرات خاصة من عصب القوقعة تستجيب لترددات خاصة مقابلة يسمى هذا التنظيم بمبدأ التوضع (Place Principle) وهو يبقى ساريا على طول الطريق السمعي إلى الدماغ. هناك مخطط آخر لتشفير المعلومات يستخدم في عملية السمع عند الإنسان يسمى مبدأ القذف (volley principle) وفيه ترسل الخلايا العصبية المعلومة عن طريق توليد نبضات كهربائية صغيرة تسمى كمون العمل. إن كل خلية عصبية على غشاء البازيلار يمكنها أن تشفر معلومة صوتية عن طريق إنتاج كمون عمل كاستجابة لكل دورة في الاهتزاز فمثلا الموجة الصوتية التي ترددها 200Hz يمكن تمثيلها بواسطة خلية عصبية واحدة تنتج 200 كمون عمل في الثانية على كل حال فإن هذا ينطبق فقط على الترددات الأقل من 500Hz وهو المعدل الأعظمي الذي تستطيع عنده الخلية العصبية إنتاج كمونات العمل. تستطيع أذن الإنسان التغلب على هذه المشكلة بالسماح لعدة خلايا عصبية أن تؤدي دورها في انجاز هذه المهمة فمثلا يمكن تمثيل النغمة التي ترددها 3000HZ بواسطة عشر خلايا تعمل بالتناوب 300 مرة في الثانية الواحدة يمدد هذا مجال مبدأ القذف إلى حوالي 4KHz.

يبين الجدول (١-١) العلاقة بين شدة الصوت وجهارته (ارتفاعه). لقد شاع التعبير عن شدة الصوت على مقياس لوغاريتمي يسمى مستوى طاقة الصوت (SPL (Sound Power Level وعلى هذا المقياس يكون الـ 0 dB SPL يمثل موجة صوتية باستطاعة مقدارها $10^{-16} Watt/Cm^2$ وهو قريب من أضعف صوت يمكن أن يسمعه الإنسان أما الكلام الطبيعي فيكون حوالي 60 dB SPL بينما الأصوات التي شدتها حوالي 140dB SPL فهي أصوات تسبب الألم للأذن [4].

الجدول (١-١) العلاقة بين شدة الصوت وجهارته

	Watts/cm^2	Decibels SPL	مثال صوتي
	10^{-2}	140 dB	مؤلم
	10^{-3}	130 dB	
	10^{-4}	120 dB	مزعج
	10^{-5}	110 dB	موسيقا الروك
	10^{-6}	100 dB	
أعلى	10^{-7}	90 dB	حدود الضجيج الصناعي
	10^{-8}	80 dB	
	10^{-9}	70 dB	
	10^{-10}	60 dB	الحديث الطبيعي
	10^{-11}	50 dB	
اخفض	10^{-12}	40 dB	اضعف صوت عند التردد 100Hz
	10^{-13}	30 dB	
	10^{-14}	20 dB	اضعف صوت عند التردد 10kHz
	10^{-15}	10 dB	
	10^{-16}	0 dB	اضعف صوت عند التردد 3kHz
	10^{-17}	-10 dB	
	10^{-18}	-20 dB	

إذاً يكون الفرق بين أعلى واخفض صوت يمكن أن تسمعه أذن الإنسان هو حوالي 120 dB وهذا يقابل بالمقياس الخطي مجالا حوالي 1 مليون ولكي تكتشف الأذن التغير بارتفاع الصوت يجب أن يكون هذا التغير بالإشارة ليس اقل من 1dB (تغير مقداره 12% بالمقياس الخطي) بعبارة أخرى يمكن القول أنه هناك فقط 120 مستوى لقياس ارتفاع الصوت تبدأ من الهمس الخافت حتى أقوى رعد.

إن حساسية الأذن مذهلة. تصور أنه عند الإصغاء للأصوات الخافتة تهتز طبلة الأذن بأقل من قطر الذرة الواحدة. يتعلق الإحساس بارتفاع الصوت بشكل تقريبي باستطاعة الصوت مرفوعة للقوة (1/3) فمثلا ارتفاع استطاعة الصوت بعامل مقداره 10 سيسجل المستمع صوتا ارتفاعه أعلى بمرتين فقط وهنا تكمن المشكلة الأساسية في التخلص من الأصوات المحيطة المزعجة كصوت ستريو عال في السكن المجاور لنفرض أنه تم تغطية 99% من الجدران بمادة عازلة للصوت وباستثناء 1% من المساحة نتيجة لوجود الأبواب و الفتحات

فبالرغم من أنه تم تخفيض استطاعة الصوت إلى 1% من قيمته السابقة فان ارتفاع الصوت قد هبط فقط إلى $0.2 \approx 0.01^{1/3}$ أو ما يقابل 20%.

بشكل عام يعتبر مجال السمع عند الإنسان من20Hz حتى 20kHz و لكنه أكثر حساسية للأصوات الواقعة في المجال من 1kHz حتى 4kHz فمثلا من الممكن كشف أصوات خافتة حتى 1dB SPL عند التردد 3 kHz و لكن عند التردد 100 Hz يجب أن تكون شدة الصوت حوالي 40 dB SPL (أي أكبر بـ 100 مرة في المقياس الخطي). أيضا يمكن للإنسان أن يميز بين نغمتين إذا كان الاختلاف بالتردد هو 0.3% عند التردد 3 kHz ولكن عند التردد 100 Hz يجب أن يكون هذا الاختلاف حوالي 3%. وللذكر فقط إن المفاتيح المتجاورة في البيانو تختلف عن بعضها بالتردد بمقدار 6% الفائدة المباشرة من وجود أذنين هي إمكانية معرفة اتجاه الصوت فالإنسان قادر على تمييز الاختلاف في الاتجاه بين منبعي صوت تفصلهما زاوية مقدارها 3 درجات ويبعدان عنه مسافة 10m. يتم الحصول على معلومات الاتجاه بطريقتين : الأولى تعتمد على تشكيل ظل قوي من قبل الرأس على الترددات الأعلى من 1 kHz بعبارة أخرى أن الأذن الأقرب إلى مصدر الصوت تلتقط إشارة أقوى من الأذن الواقعة على الطرف الأخر من الرأس. أما الطريقة الثانية فتعتمد على تأخير الصوت قليلا إلى الأذن الواقعة على الطرف البعيد للرأس بالنسبة للإذن القريبة و ذلك نتيجة لبعدها عن منبع الصوت وبالأخذ بعين الاعتبار القياس النموذجي للرأس (حوالي 22 Cm) و سرعة الأمواج الصوتية (340 متر في الثانية) فان الحصول على قدرة تميز بالزاوية مقدارها 3 درجات يتطلب ضبط زمني حوالي 30 ميكرو ثانية. بما أن هذا التوقيت يتطلب مبدأ القذف فإن هذه الطريقة للتوجيه تستخدم بشكل أساسي للترددات الأقل من 1 kHz.

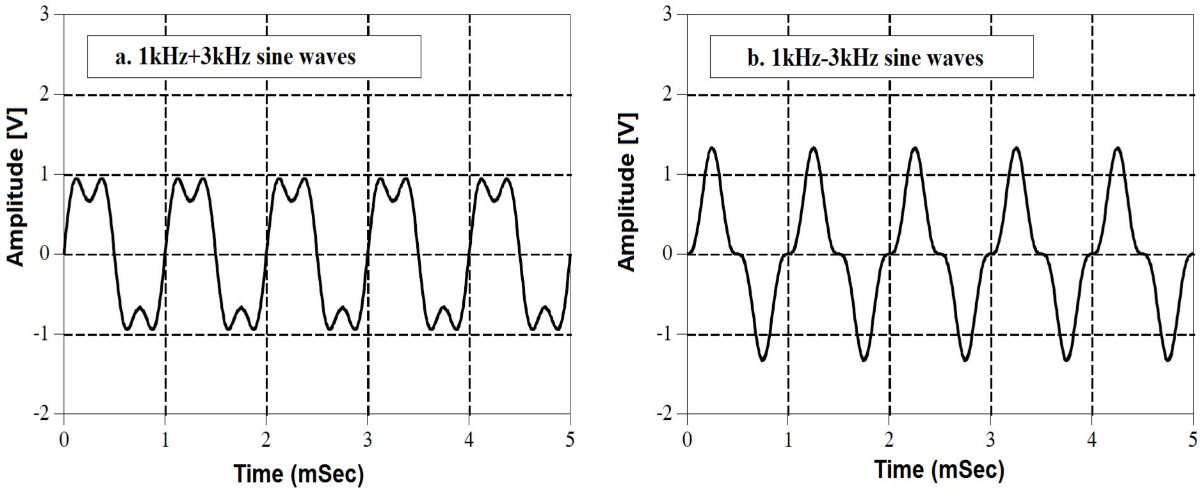
يتم دعم هذان المنبعان في الحصول على معلومة الاتجاه إمكانية لفت الرأس و مراقبة التغير بالإشارات. يشعر الإنسان بإحساس ممتع عندما يصل إلى كلتا الأذنين نفس الصوت كالاستماع إلى صوت وحيد البعد (ليس ستريو) بواسطة سماعات رأسية حيث سيشكل انطباع في الدماغ أن الصوت آت من مركز الرأس.

صحيح إن الإنسان قادر على تحديد اتجاه الصوت إلا أن قدرته على تحديد مسافات الصوت ضعيفة وذلك لأن موجة الصوت تحوي القليل عن هذه المعلومة. اجل فالإنسان بالكاد يميز أن أصوات الترددات العالية تكون قريبة بينما أصوات الترددات المنخفضة تكون بعيدة والسبب في ذلك أن الموجة الصوتية تبدد تردداتها العالية عند انتشارها لمسافات بعيدة أما ظاهرة الصدى فهي تشكل انطباعا عن حجم الغرفة ولا يمكن الاعتماد عليها في تحديد المسافة فمثلا إن الأصوات في القاعة الكبيرة ستحوي صدى بفاصل زمني حوالي 100 msec بينما يكون هذا الفاصل حوالي 10msec في المكتب الصغير. هناك بعض المخلوقات التي حلت مشكلة معرفة المسافة عن طريق استخدام مبدأ السونار الايجابي (Active Sonar)كالدلافين التي تصدر طقطقة وصريرا تنعكس عن الأجسام القريبة وبقياس الفاصل الزمني بين لحظة إصدار الصوت وارتداد الصدى يمكن لهذه المخلوقات إن تحدد توضع الأجسام بقدرة تصل إلى 1Cm. لقد بينت التجارب أن بعض الأشخاص خاصة العميان قادرون على استخدام الصدى لتحديد الموقع لمسافات قريبة [4].

١-٤-٢ طابع الصوت

غالبا يتم تقسيم عملية الإحساس بالأصوات المستمرة كالنغمة الصادرة عن أداة موسيقية إلى ثلاثة أقسام: جهارة (ارتفاع) الصوت (loudness) ونغمته (Pitch) وطابعه (timbre). يعتبر ارتفاع الصوت مقياسا لشدته كما سبق وشرحنا بينما النغمة هي المركبة الترددية الأساسية في الصوت أي التردد التكراري لموجة الصوت وفي حين إنه هناك تأثيرات دقيقة لهذين الأحساسين فهما مرتبطان بشكل صريح مع كميات فيزيائية يمكن تمييزها بسهولة.

يعتبر الطابع إحساسا أكثر تعقيدا فهو يتحدد بالمكونات التوافقية (مضاعفات التردد الأساسي) للإشارة يوضح الشكل(١-١٤) موجتين متشكلتين عن طريق جمع موجة جيبية ترددها 1kHz ومطالها واحد مع موجة جيبية أخرى ترددها 3 kHz ومطالها يساوي النصف نلاحظ أن الموجتين مختلفتين بالشكل وهذا سببه عكس الصفحة لمركبة التردد العالي قبل عملية الجمع في الموجة الظاهرة في الشكل (b-١٤-١) أي إن التوافقية الثالثة (3 kHz) قد تعرضت لإزاحة صفحية مقدارها 180 درجة بالنسبة للتوافقية الأولى (1 kHz) وبالرغم من الاختلاف الكبير بين الأمواج في مجال الزمن فإنهما متماثلتان بالصوت وهذا تفسيره أن حاسة السمع مبنية على مطال الترددات وهي غير حساسة لصفحتها. هذا يعني إن علاقة شكل موجة الصوت في مجال الزمن مع حاسة السمع هو ارتباط غير مباشر وغالبا لا يؤخذ بعين الاعتبار في المنظومات الصوتية.

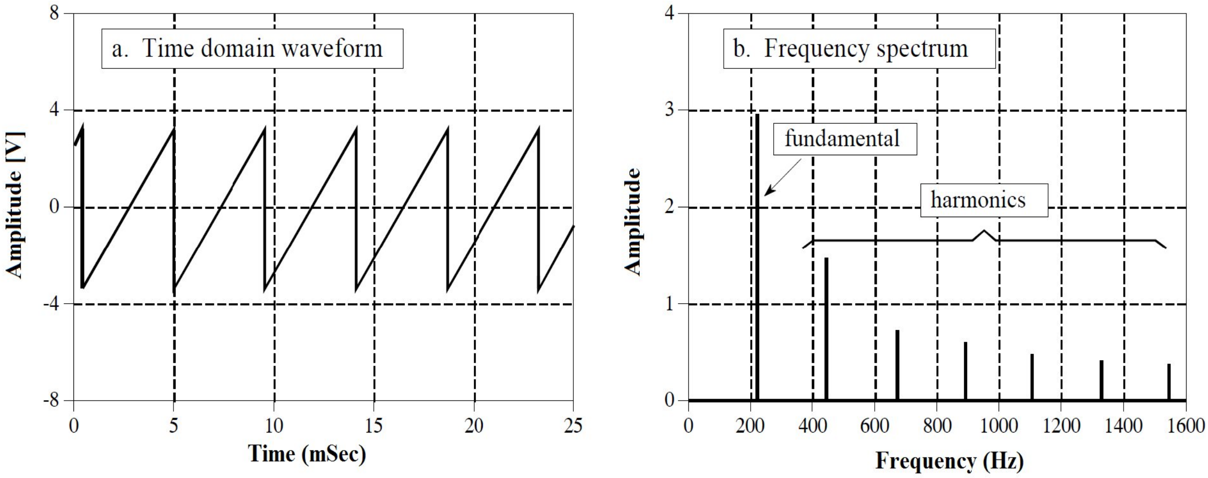


الشكل (١-١٤) يوضح مفهوم طابع الصوت

يمكن فهم عدم حساسية الأذن للصفحة عن طريق دراسة كيفية انتشار الصوت في الوسط المحيط. لنفرض أنك تستمع لشخص ما يحدثك عبر غرفة صغيرة فإن معظم الصوت يصل إليك منعكسا عن الجدران والسقف والأرض وبما أن انتشار الصوت يعتمد على التردد فإنه سيصل إلى أذنيك ترددات مختلفة سالكة مسارات مختلفة هذا يعني إن الصفحة النسبية لكل تردد ستتغير عندما تتجول حول الغرفة ولأن الأذن تهمل هذا التغير الصفحي فإنه خلال التجول لن تشعر بأي تغير في الصوت ومن وجهة نظر فيزيائية تصبح صفحة موجة الصوت ذات طبيعة عشوائية خلال الانتشار و الأذن تهملها لأنها لا تحمل معلومات مفيدة. على كل حال لا يمكن القول إن الأذن صماء بشكل كامل بالنسبة للصفحة و ذلك لأن تغير الصفحة يمكن أن يعيد التسلسل الزمني لإشارة الصوت

كمثال على هذا نذكر منظومة chirp التي تقوم بتحويل نبضة قصيرة إلى إشارة أطول بكثير فبالرغم من أنهما مختلفتين فقط بالصفحة فإن الأذن قادرة على التميز بين الصوتين بسبب اختلاف مدتهما.

لنفرض أننا نطلب من عازف كمان أن يعزف علامة ما لتكن العلامة A الواقعة قبل علامة C الوسطى عندما تظهر الموجة على راسم الإشارة ستبدو كموجة سن المنشار المبينة في الشكل (١–١٥–a) وما هذا إلا نتيجة للقلفونه الموجودة على شعيرات قوس الكمان. عندما يجر العازف القوس على الوتر ستتشكل الموجة طالما أن القوس ملتصق بالوتر وبرفع القوس ستنقطع نهائيا تكرر هذه العملية نفسها مرة تلو مرة معطية بشكل سن المنشار. يبين الشكل (١–١٥–b) كيف تحس الأذن بهذا الصوت فكما هو ملاحظ إنه مكون من تردد أساسي 220 Hz وتوافقياته عند الترددات 440Hz 660Hz 880 Hz... إذا تم عزف نفس العلامة بواسطة آلة موسيقية أخرى سيكون شكل الموجة مختلفا ومع ذلك ستبقى الأذن تسمع التردد 220 Hz وتوافقياته بما أن الأداتين تعطيان نفس التردد الأساسي لهذه العلاقة فالصوتان سيكونان متشابهين ويقال إن الأداتين تملكان نفس النغمة وحيث إن المطال النسبي للتوافقيات مختلف لن يكون الصوتان متماثلان ويقال إن الأداتين لهما طابع مختلف.



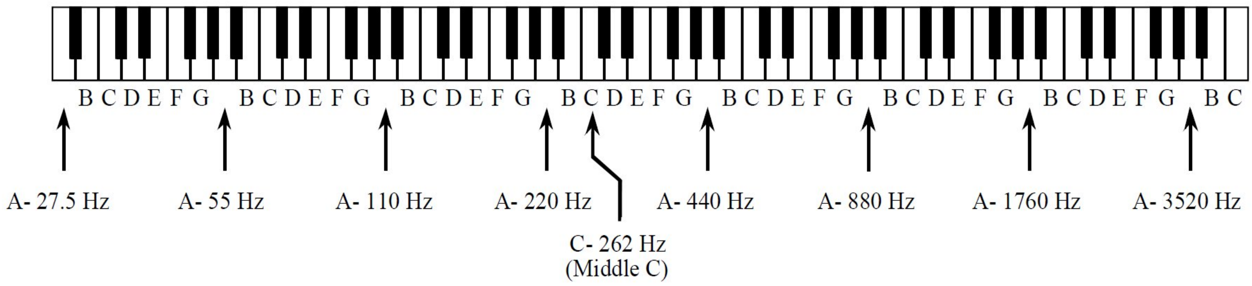
الشكل (١–١٥) التوافقيات الأساسية لإشارة سن منشار

غالبا ما يقال إن طابع الصوت يتحدد بشكل الموجة، إن هذا صحيح ولكن قد يكون مضللا بعض الشيء. ينتج الإحساس بالطابع من كشف الأذن التوافقيات، وبينما يتحدد محتوى الموجة من التوافقيات بشكل هذه الموجة فإن عدم حساسية الأذن للصفحة تجعل العلاقة وحيدة الجانب أي إن موجة محددة يمكن أن تملك جرسا واحداً في حين إن جرسا ما يمكن إن يكون له عدد لانهائي من الأمواج.

إن الأذن معتادة على سماع التردد الأساسي وتوافقياته فإذا قدمنا لمستمع ما تركيبة من موجتين جيبيتين ترددهما 1kHz و3kHz سيعطيان صوتا طبيعيا وممتعا أما إذا استخدمنا موجتين بتردد 1 kHz و3.1kHz فالصوت الناتج سيكون نشازاً (سيكون الصوت في هذه الحالة مختلفاً عن الصوت في حالة الترددين 1kHz و3kHz).

نوضح في الشكل(١–١٦) السلم الموسيقي الأساسي من خلال مفاتيح البيانو. يعطي الضرب على المفتاح الموجود في الطرف اليساري للبيانو التردد الأساسي وقدره 27.5Hz بالإضافة إلى توافقياته عند الترددات 55Hz 110Hz 220Hz 440Hz 880Hz الخ (هناك أيضا توافقيات أخرى بين هذه الترددات ولكنها ليست

ذات أهمية في مناقشتنا هنا) تقابل هذه التوافقيات الترددات الأساسية التي تعطيها مفاتيح أخرى في اللوحة و بالتحديد فان كل سابع مفتاح ابيض هو احد توافقيات المفتاح اليساري أي أن المفتاح الثامن من اليسار ذو تردد أساسي 55Hz والمفتاح الخامس عشر ذو تردد أساسي 110Hz الخ وكون هذه المفاتيح توافقيات لبعضها البعض فإنها تعطي أصواتًا متماثلة و تكون متناغمة عند الضغط عليها بالتوافق، لهذا السبب تسمى جميعها بالعلامة A بنفس الشكل فان المفاتيح التي تأتي مباشرة عن يمين المفاتيح A تسمى المفتاح B و جميعها تكون متناغمة مع بعضها.



الشكل (١-١٦) السلم الموسيقي

يتكرر نفس هذا النمط من أجل العلامات الموسيقية السبع A,B,C,D,E,F,G. المصطلح أوكتاف octave في الموسيقى يعني عاملا مقداره 2 بالنسبة للتردد ففي البيانو أوكتاف واحد يشمل ثمانية مفاتيح بيض وهذا نسبة للاسم octo باللاتيني والذي يعني ثمانية؛ بعبارة أخرى إن ترددات البيانو تتضاعف كل ثمانية مفاتيح بيض ولوحة المفاتيح تمتد على سبعة أوكتافات أما مجال السمع عند الإنسان فقد اصطلح عليه من 20 Hz حتى 20 kHz وهذا يقابل 1/2 أوكتاف إلى اليسار و2 أوكتاف إلى اليمين وبما أن الأوكتافات مبنية على مضاعفة التردد بعد عدد محدد من المفاتيح فيمكن اعتبارها تمثيلا لوغاريتميا للتردد وهذا أمر هام لأن معلومات الصوت بشكل عام يتم توزيعها وفق هذه الطريقة، فعلى سبيل المثال إن كمية المعلومات الموجود بين الاوكتافين 50Hz و 100Hz هي نفس كمية المعلومات الموجودة بين الاوكتافين 10 kHz و20 kHz وبالرغم من أن البيانو يغطي فقط 20% من مجال الترددات المسموعة عند الإنسان (4kHz من أصل 20 kHz) فإنه قادر على أن يعطي أكثر من 70% من المعلومات الصوتية التي تفهمها الأذن (7 أوكتافات من أصل 10). بنفس الشكل فإن الترددات العليا التي يمكن أن يسمعها الإنسان تضيع مع تقدم العمر و هي حوالي 20 kHz إلى 10 kHz وهذا لا يشكل إلا 10% من إمكانية السمع (أوكتاف واحد من أصل 10). [4].

١-٤-٣ جودة الصوت ومعدل المعطيات

عند تصميم منظومة صوت رقمية يجب طرح سؤالين:(1) ما مقدار الحاجة لأن يكون الصوت جيدا (2) ما هو معدل المعطيات الذي يمكن السماح به ؟ إن الجواب على هذين السؤالين يقودنا للتميز بين ثلاث فئات للصوت: الأولى وهي الموسيقى العالية الأمانة حيث تكون جودة الصوت ذات أهمية كبيرة و على الأغلب سيكون معدل المعطيات مقبولا؛ الثانية هي فئة الاتصالات الهاتفية وتتطلب تأمين الكلام بصوت طبيعي ومعدل معطيات منخفض من أجل خفض كلفة المنظومة؛ الثالثة هي فئة الكلام المضغوط حيث يكون تخفيض معدل المعطيات

ذات أهمية كبيرة مع إمكانية التسامح بانخفاض جودة الصوت وهذا يشمل الاتصالات العسكرية والهاتف النقال والكلام المخزن رقميا من أجل البريد الصوتي والملتيميديا.

في منظومات الموسيقى العالية الأمانة يكون التقطيع سريعا بما فيه الكفاية (44.1kHz) وأيضا تستخدم دقة كافية (16 bit) لكل عينة بحيث يمكنها أن تلتقط جميع الأصوات التي يمكن أن يسمعها الإنسان. إن هذه الجودة العظيمة في الصوت تأتي على حساب معدل معطيات عال: 44.1kHx16bit=706Kbit/Sec في حين إن الموسيقى تتطلب عرض حزمة حوالي 20kHZ فإن الكلام العادي لا يحتاج إلى أكثر من 3.2kHz وبالرغم من إن المجال الترددي قد نقص إلى 16% فقط (3.2kHz من أصل 20 kHz) فان الإشارة ما تزال تحوي على 80% من معلومات الصوت الأصلي (8 أوكتافات من أصل 10). تعمل منظومات الاتصالات بشكل نموذجي بمعدل تقطيع حوالي 8 kHz وهذا يسمح بالاحتفاظ بالصوت الطبيعي ولكن جودة الموسيقى تنقص بشكل كبير.

لذا لابد أن نعرف جودة الصوت: إن محطة الراديو FM تبث على حزمة عرضها 20 kHz بينما محطات الراديو AM تكون عصبتها محدودة بـ 3.2kHz فالأصوات تسمع بشكل طبيعي على المحطات AM إلا إن الموسيقى تكون ضعيفة وغير مرضية. أيضا في منظومات الصوت فقط يتم خفض الدقة من 16 bit إلى 12 bit لكل عينة مما يغير جودة الصوت بشكل بسيط ويمكن أيضا أن تخفض إلى 8 bit لكل عينة إذا كانت خطوة التكميم غير متساوية يسمى هذا الإجراء الواسع الانتشار بتقنية ضغط بسط(companding).

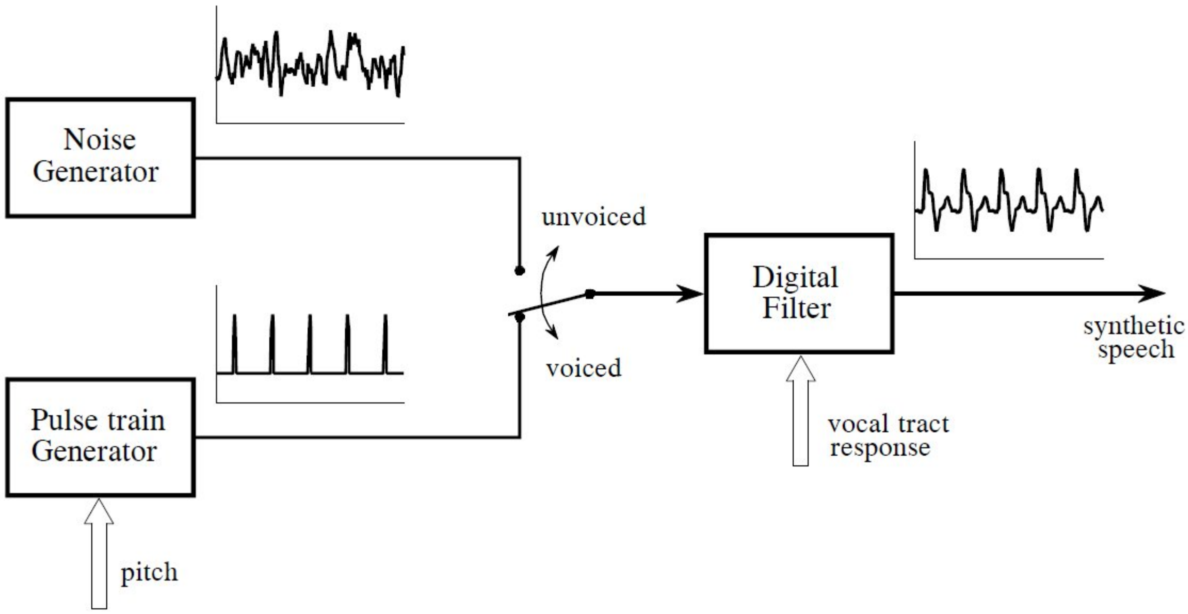
إن معدل تقطيع مقداره 8kHz واستخدام ADC ذي دقة 8 bit لكل عينة يعطي معدل معطيات 64 Kbit/sec ومن أجل الكلام العادي يعتبر هذا معدلا قويا.

يمثل معدل المعطيات 64 Kbit/sec تطبيقا مباشرا لنظرتي أخذ العينات و التكميم على إشارات الصوت. هناك تقنيات مستخدمة لتخفيض معدل المعطيات مبنية على ضغط سلاسل المعطيات عن طريق حذف الفائض في إشارات الكلام [4].

١-٤-٤ تمييز الصوت

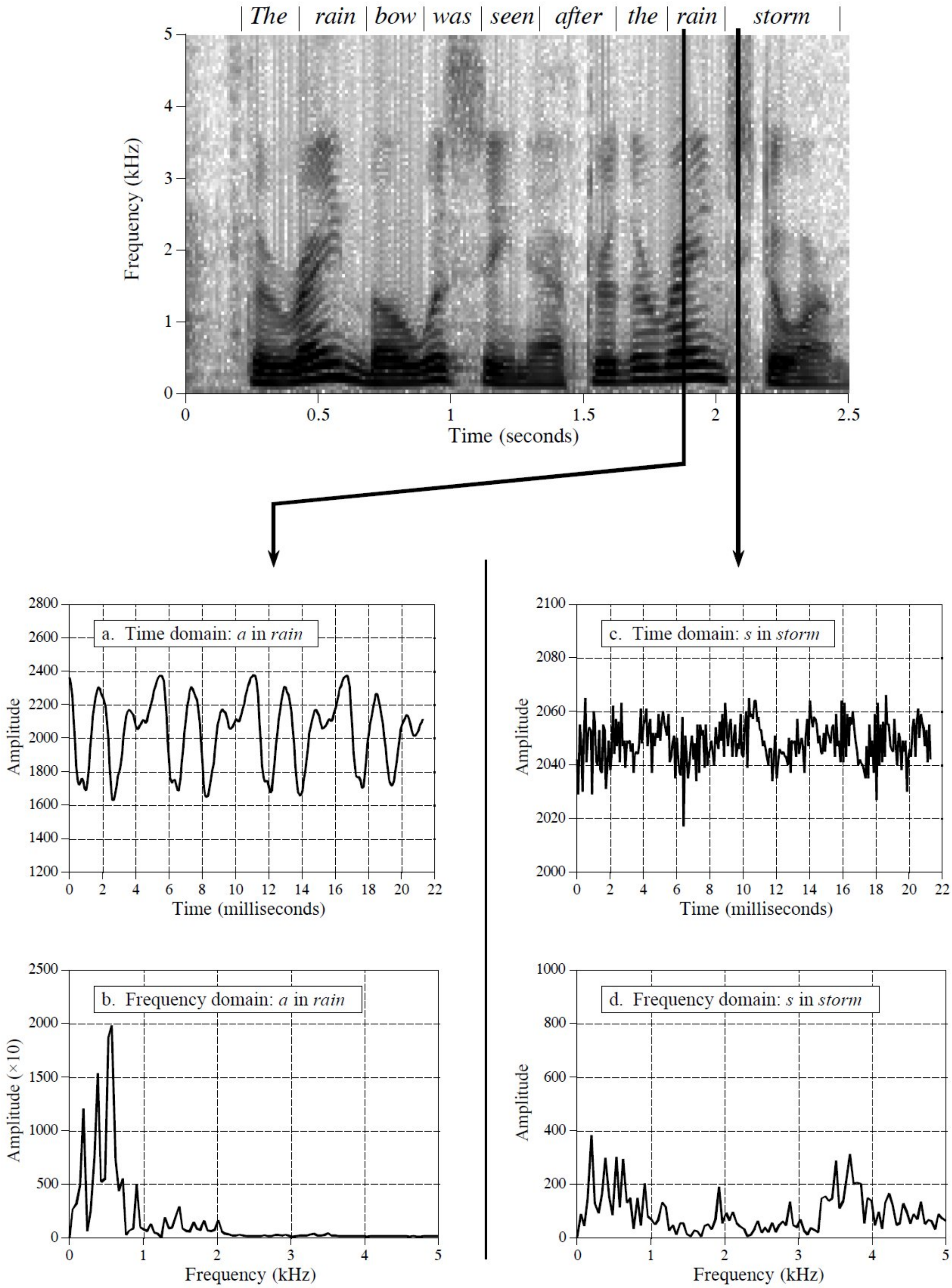
إن استخدام الحاسب لتوليد الكلام والتعرف عليه من المسائل المرعبة والصعبة فهناك العديد من الأساليب التي طرقت هذه المسألة ولكن النجاحات كانت قليلة؛ فهذا المجال يعتبر من مجالات البحث الفعالة في معالجة الإشارة الرقمية. سنقدم هنا فقط مقدمة موجزة عن الطرق النموذجية التي استخدمت في هذا المجال.

إن أغلب منظومات تركيب الكلام والتعرف عليه تبنى على مخطط إنتاج الكلام المبين في الشكل (١-١٧). يعتمد هذا المخطط في عمله على تصنيف الحروف التي ينطقها الإنسان إلى فئتين أحرف صوتية (voiced)وأحرف الصفير (fricative). تتكون الأحرف الصوتية عندما يخرج الهواء من الرئتين ويمر على الحبال الصوتية ومن ثم يخرج من الفم و/أو الأنف. الحبال الصوتية عبارة عن شريطين من نسيج رقيق ممدودين عبر مجرى الهواء لتفاحة آدم (الحنجرة) وكاستجابة لتغير توتر العضلة تهتز الحبال الصوتية بترددات بين 40 Hz و 600 Hz وينتج عن هذا نفخات دورية تحقن إلى الحنجرة. الأحرف اللينة هي مثال على هذه الأصوات. في الشكل (١-١٧) تم تمثيل عملية توليد الأحرف الصوتية بمولد قطار نبضات مع معيرة لتحديد النغم (pitch) (أي التردد الأساسي للموجة).



الشكل (١-١٧) مخطط إنتاج الكلام

بالمقابل فإن أحرف الصفير تنشأ على شكل ضجيج عشوائي ليس من اهتزاز الحبال الصوتية وهذا يحدث عندما يمر الهواء بالقرب من اللسان أو الشفاء أو/و الأسنان مما يؤدي إلى اضطراب الهواء قرب الاختناق. تشمل أحرف الصفير الحروف التالية S, F, Sh, z, v, Th. في الشكل (١-١٧) مثلنا عملية توليد أحرف الصفير بمنبع ضجيج.إن كلا من الأحرف الصوتية وأحرف الصفير يتم تشكيلها في التجاويف الصوتية المكونة من اللسان والشفاه والفم والحنجرة والمجرى الأنفي وبما أن مرور الصوت عبر هذه البنية هي عملية خطية فيمكن تمثيلها بمرشح خطي ذي استجابة نبضية مناسبة؛ وفي اغلب الحالات يستخدم في هذا المخطط مرشح تراجعي وتكون خواصه محددة بالثوابت التراجعية ولأن إبعاد التجويف الصوتي تبلغ عدة سنتيمترات تكون الاستجابة الترددية سلسلة من الترددات الطنينية في مجال الكيلو هرتز؛ وبلغة معالج الأصوات تسمى قمم هذه الترددات الطنينية بترددات الشكل (format frequency). بتغير الوضعية النسبية للشفاه واللسان يمكن أن تتغير ترددات الشكل بالمطال والتردد.يبين الشكل(١-١٨) طريقة شائعة في تمثيل إشارات الكلام عن طريق الصورة الطيفية (spectrogram) أو طبعة الصوت(voiceprint) تقسم إشارة الصوت إلى أجزاء صغيرة. لنفرض أن طولها من 2 msec إلى 40 msec ومن ثم يستخدم FFT لإيجاد الطيف الترددي لكل جزء؛ توضع الأطياف الناتجة بجانب بعضها البعض وتحول بعدها إلى صورة بتدرج رمادي (يقابل المطال المنخفض لون فاتح والمطال العالي يقابله لون عاتم). إن هذا يعطي طريقة لمراقبة تغيرات المركبات الترددية للكلام مع الزمن أما طول كل جزء فيتم اختياره بالتوفيق بين دقة التردد (تفضل أن تكون الأجزاء طويلة) ودقة الزمن (تفضل إن تكون الأجزاء قصيرة). نلاحظ من منحني الحرف a في كلمة rain الشكل(١-١٨-a) إن الأحرف الصوتية تكون في مجال الزمن على شكل موجة دورية ويكون طيفها الترددي على شكل مجموعة من التوافقيات الموزعة بانتظام الشكل(١-١٨-b).



الشكل(١-١٨) تحليل الكلام عن طريق الصورة الطيفية

بالمقابل يبين الحرف s في كلمة Storm إن أحرف الصفير تملك طابعا ضجيجيا في مجال الزمن الشكل(١-١٨-c) وأيضا يكون طيفها الترددي عشوائيا الشكل(١-١٨-d) تبين هذه الأطياف أيضاً الشكل بواسطة ترددات التشكيل. من أجل كلا الصوتين نلاحظ أيضا على الصورة التي تظهر العلاقة تردد- زمن. إن مظهر الحرف a في كلمة rain الأولى مشابه لمظهره في الكلمة الثانية.

يمكن تقريب الإشارة الكلامية خلال زمن قصير لنقل 25 msec عن طريق تحديد البارمترات الثلاثة التالية:

(1) الاختيار بين تهييج دوري أو عشوائي(حرف صوتي أو حرف صفير) (2) تردد الموجة الدورية (في حال اختيارها) (3) اختيار ثوابت المرشح الرقمي الذي تم اختياره لتقليد المسار الصوتي أما الكلام المستمر فيكون من الممكن تركيبه عن طريق التحديث المستمر لهذه البارمترات 40 مرة في الثانية. لقد كانت هذه الطريقة وراء واحد من النجاحات التجارية المبكرة في DSP ألا وهو المخابر الالكترونية لتعليم الأطفال تكلم وتهجى (Speak & Spell) والتي انتشرت بكثرة. لقد كانت جودة الصوت ضعيفة في مثل هذه النماذج لتركيب الكلام فكان الصوت ميكانيكياً ومختلفاً عن صوت الإنسان ومع ذلك فهو يتطلب معدل معطيات منخفض بشكل نموذجي حوالي بضع كيلوبت في الثانية.

أيضا كان هذا هو الأساس المعتمد في طريقة الترميز بالتنبؤ الخطي (LPC) (Codinglinear Predictive) لضغط الكلام [4]. في هذه الطريقة تقسم التسجيلات الكلامية الرقمية إلى أجزاء قصيرة يملك كل منها ميزاته الخاصة وفقا البارامترات الثلاثة للنموذج يتطلب هذا حوالي 2 Kbytes/sec إلى 6 Kbytes/sec. المعلومة الموجودة في كل جزء يمكن أن ترسل أو تحفظ حسب الحاجة ومن ثم يعاد بناؤها بمركب الكلام.

أما خوارزميات التعرف على الكلام الهدف منها هي محاولة التعرف على النماذج من البارمترات المستنتجة. يتضمن هذا مقارنة المعلومة الموجودة في الجزء مع قوالب صوتية مخزنة مسبقا لمحاولة التعرف على الكلمة التي نطقت. هنا توجد مشكلة فهذه الطريقة لا تعمل كما يجب و هي تفيد في بعض التطبيقات فقط و لكن إمكانياتها تبقى دون إمكانيات السمع عند الإنسان. لفهم صعوبة التعرف على الكلام نتصور أن أحدا ما غير متوقع ينطق الجملة التالية:

Larger run medical buy dogs fortunate almost when

من الطبيعي أن لا نفهم معنى لهذه الجملة لأنها لا تحمل أصلا أي معنى والأهم من ذلك أنه من المحتمل إننا لن نفهم بعض الكلمات الفردية، إن هذا أساسي بالنسبة للطريقة التي يعتمدها الإنسان لفهم الكلام فالكلمات تعرف من أصواتها وأيضا من سياق الكلام ومن توقع المستمع، تصور على سبيل المثال إننا نسمع العبارتين التاليتين:

The child wore a spider ring on Halloween

He was an American spy during the war

حتى ولو أن الكلمات المسطرة في الجملتين كان لها نفس الصوت فالمستمع سيسمع الكلمات الصحيحة وذلك من سياق الجملة. إن هذا ليس متعمدا فعلا بل إنه جزء متأصل في عملية السمع عند الإنسان.

إن اغلب خوارزميات التعرف على الكلام تثق بأصوات الكلمات الفردية وليس بسياق الكلام فهي تحاول التعرف على الكلمات وليس فهم الكلام. إن هذا الوضع يضعها في مكانة متدنية جدا بالمقارنة مع السمع عند الإنسان.

بشكل عام تواجه منظومات التعرف على الكلام ثلاث عقبات أساسية وهي: (1) لا بد من وجود توقفات بين

الكلمات المطلوب التعرف عليها؛ إن هذا يلغي حاجة الخوارزمية للتعامل مع العبارات متشابهة الصوت والمكونة من كلمات مختلفة (مثل Spider ring و Spy during).لا تسبب هذه الحالة البطء والإرباك للأشخاص المعتادين على الكلام بسرعة (2) إن عدد المفردات اللغوية محدود فقط بعدة مئات من الكلمات وهذا يعني أنه على الخوارزمية أن تبحث في مجموعة محدودة لتجد الكلمات الأكثر توافقا؛ كما أن زيادة عدد المفردات ستؤدي إلى زيادة زمن التعرف وأيضا معدل الخطأ (3) يجب تدريب المنظومة على أصوات جميع الأشخاص الذين سيستخدمونها وهذا يتطلب أن يقوم كل شخص من هؤلاء بنطق كل كلمة يطلب التعرف عليها.

غالبا سيحتاج هذا إلى التكرار من خمس إلى عشر مرات؛ إن هذا الطابع الشخصي لقاعدة المعطيات يزيد من دقة التعرف على الكلمات إلا أن هذا غير مناسب ومبدد للوقت. أن تطوير تقنية ناجحة للتعرف على الكلام عملية هائلة التكلفة ويعتبر الكلام الطريقة الأسرع والأجدى للتواصل بين البشر ومن شأن التعرف على الكلام إن يحل بديلا عن الكتابة و عن الإدخال بواسطة لوحة المفاتيح وعن دارات التحكم الالكترونية التي تستخدم المفاتيح والأزرار. يكفي فقط أن يصبح عملها أفضل قليلا لكي تروج في السوق التجارية إن التطور المنظور في التعرف على الكلام سيأتي نتيجة لتطور مماثل في مجال الذكاء الصنعي والشبكات العصبونية [4].

١-٤-٥ المعالجة غير الخطية للصوت

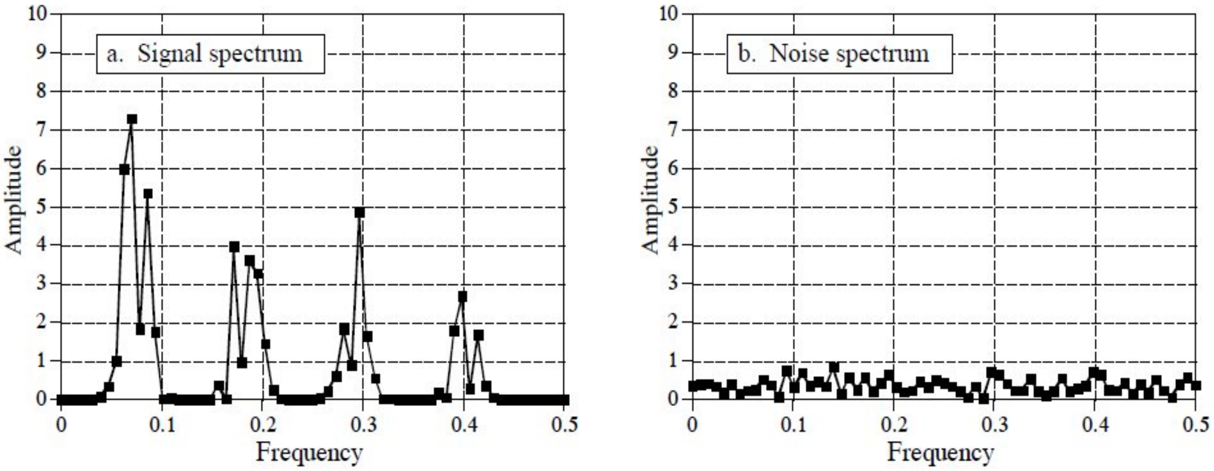
إن المعالجة الرقمية قادرة على تحسين إشارات الصوت بطرق عديدة فعلى سبيل المثال يمكن استخدام مرشح وينر لفصل المركبات التي تغلب فيها الإشارة عن المركبات التي يغلب فيها الضجيج. بنفس الشكل فإن التراجع عن الانطواء يمكن أن يحذف عمليات الانطواء غير المرغوب بها كعملية استرجاع تسجيلات قديمة. تشكل هذه النماذج من التقنيات الخطية العمود الفقري لـ DSP. هناك أيضا العديد من التقنيات غير الخطية التي تفيد في مجال معالجة الصوت سنصف هنا بإيجاز اثنتين من هذه التقنيات:

التقنية الأولى غير الخطية هي تقنية مخصصة لتخفيض مستوى الضجيج العريض الحزمة في إشارات الكلام مثل ضجيج الفحيح في الأشرطة الممغنطة والضجيج الالكتروني في الدارات التشابهية وصوت الريح في الميكرفون .. الخ. نادراً ما تستخدم المرشحات الخطية في مثل هذه الحالات والسبب في ذلك أن ترددات الضجيج متداخلة بشكل كامل مع ترددات إشارة الصوت فكلاهما يشغل المجال من 200 Hz إلى 3.2kHz. كيف يمكن فصل إشارتين عندما تكونان متداخلتين في مجال الزمن ومجال التردد ؟

سنبين هنا كيف تحل هذه المسألة. لو أخذنا الطيف الترددي لجزء صغير من الإشارة سنجد أن مطال المركبات الترددية مختلف بشكل كبير. سنأخذ كمثال الطيف الترددي المبين في الشكل(١-١٩-a) وهو طيف إشارة كلامية طولها 16 msec (أي ما يقابل 128 عينة عند ما يكون معدل التقطيع 8kHz) إن الجزء الأكبر من الإشارة يكون موجودا في المركبات الترددية العالية المطال بالمقابل يوضح الشكل(١-١٩-b) طيف إشارة الضجيج فقط ومن الواضح أنه طيف فوضوي إلا أنه أكثر انتظاما عند الترددات المنخفضة.

الآن سنبين الفكرة الأساسية: عند وجود الإشارة والضجيج معا فإنه من الممكن فصلهما بالنظر إلى مطال كل مركبة ترددية، فإذا كان المطال كبيرا فإنه على الأغلب ينتمي للإشارة ومن الطبيعي إن نحافظ عليه أما إذا كان

المطال صغيرا فمن الممكن أن ينسب للضجيج وبالتالي سيتم إهماله أي يوضع صفرا وتضبط المركبات الترددية المتوسطة بطريقة ما بين النهايتين.

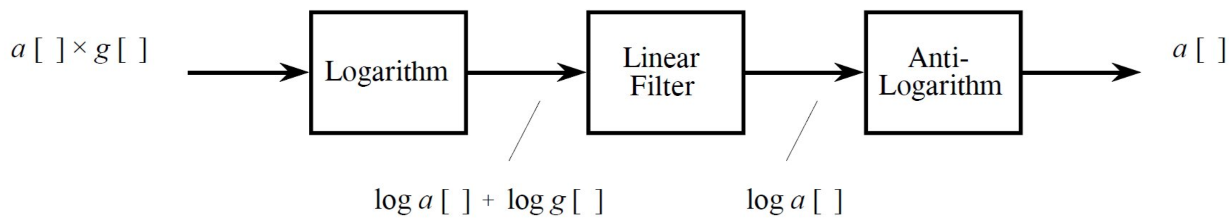


الشكل (١-١٩) الطيف الترددي مع الضجيج

سنعرض هذه التقنية بطريقة أخرى على شكل مرشح وينر المتغير مع الزمن فكما نذكر أن الاستجابة الترددية لمرشح وينر تمرر الترددات التي تكون فيها الإشارة غالبة وتحذف الترددات التي يكون فيها الضجيج غالبا؛ يتطلب هذا المعرفة المسبقة لطيف الإشارة وطيف الضجيج. تستخدم هذه التقنية غير الخطية نفس الفكرة ماعدا أنه يجب إعادة حساب الاستجابة الترددية لمرشح وينر من أجل كل جزء من أجزاء الإشارة وذلك بالاعتماد على طيف ذلك الجزء. بعبارة أخرى إن الاستجابة الترددية للمرشح متغيرة من جزء إلى آخر وذلك وفقاً لمميزات الإشارة نفسها.

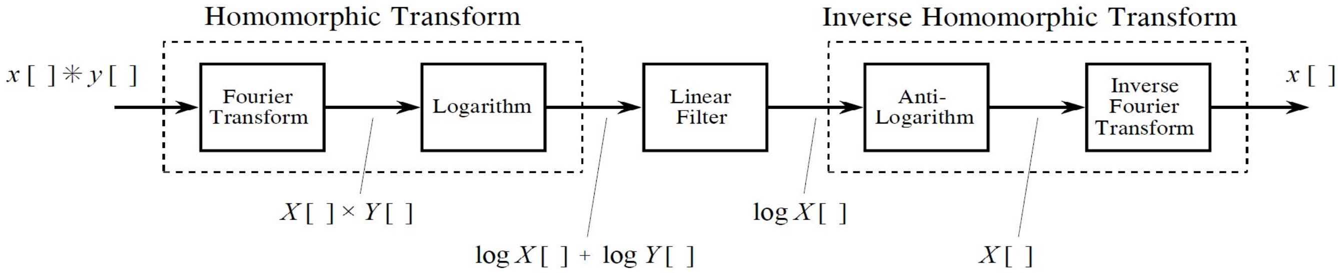
إحدى الصعوبات التي تواجه هذه التقنية (وغيرها من التقنيات) غير الخطية هي عدم صلاحية طريقة التداخل مع الجمع في ترشيح الإشارات الطويلة فبما أن الاستجابة الترددية تتغير مع الزمن، فإن شكل الموجة في كل جزء من أجزاء مجال الزمن لن يكون متناسقا مع الأجزاء المجاورة. من الممكن التغلب على هذا بالتذكير أن معلومات الصوت محتواة في النماذج الترددية المتغيرة مع الزمن وليس بشكل الموجة في مجال الزمن. هناك طريقة نموذجية تعتمد على تقسيم الإشارة في مجال الزمن إلى أجزاء متداخلة و بعد المعالجة يتم تطبيق نافذة ضيقة على الأجزاء المتداخلة قبل أن يتم ضمها إلى بعضها البعض. يؤمن هذا انتقالاً ناعماً للطيف الترددي من جزء إلى آخر.التقنية الثانية غير الخطية التي سنستعرضها هنا هي طريقة المعالجة التماثلية للإشارة (homomorphic signal processing). إن الجمع ليس الطريقة الوحيدة التي يضاف بها الضجيج و التداخل إلى الإشارة المفيدة فالجداء والانطواء هي أيضا طرق لمزج الإشارات مع بعضها فإذا تم ضم الإشارات بطريقة غير خطية (أية طريقة غير الجمع) فإنه لا يمكن فصلها باستخدام مرشحات خطية. تحاول تقنية المعالجة التماثلية فصل مثل هذه الإشارات بتحويل هذه المسالة وجعلها خطية أي إن المسألة قد تحولت إلى نفس البنية كمنظومة خطية. لنفترض على سبيل المثال أنه تم إرسال إشارة صوت على شكل موجة راديو معدلة مطاليا(AM)، ستؤدي تغيرات وسط الانتشار إلى تغير مطال الإشارة الملتقطة مما يؤدي إلى تغير ارتفاع إشارة الصوت مع الزمن. يمكن وضع نموذج لهذا الحالة على شكل إشارة صوت ممثلة بـ a[] مضروبة بإشارة متغيرة

ببطء مع الزمن g[] في الدارات الالكترونية التشابهية يتم التعامل مع مثل هذه المسالة باستخدام دارات التحكم الآلي بالربح AGC ولكن من الممكن أيضا تصحيحها بواسطة المعالجة غير الخطية في DSP.
كما هو مبين في الشكل (١-٢٠) تمرر الإشارة الكلية[]g .[ ]a على تابع لوغاريتمي ومن المساواة log (x.y) = logx+logy يمكن تحويل العملية غير الخطية (الجداء) إلى عملية خطية (جمع) أي سنحصل على[]loga[]+logg. بعبارة أخرى أن اللوغاريتم هو لتحويل التماثلي الذي يعيد المسألة غير الخطية إلى مسألة خطية. بعد ذلك يتم فصل الإشارات التي كونت الجمع بواسطة مرشح خطي مناسب أي أن بعض الترددات ستمر بينما بعضها الأخر سيحذف. في حالة الـ AGC ستكون إشارة الربح g[] مكونة من ترددات منخفضة أقل بكثير من المجال 200 Hz إلى 3.2kHz الموافق للإشارة الصوتية. إن طيف لوغاريتم هذه الإشارات سيكون معقدا ولكن هذا لا يغير من الفكرة. يستخدم مرشح تمرير عالي لحذف مركبات تغيرات الربح من الإشارة بالنتيجة تتحول الإشارة []loga[]+logg إلى []loga وفي الخطوة الأخيرة يتم التراجع عن اللوغاريتم باستخدام تابع أسي (عكس اللوغاريتم أو ex) فنحصل على إشارة الخرج المطلوبة []a.
يبين الشكل (١-٢١) منظومة تماثلية لفصل إشارات كانت قد خضعت للانطواء وقد أثبتت هذه المنظومة فائدة في عملية تخليص إشارات الصوت من الصدى، أي إن إشارة الصوت قد خضعت لانطواء مع استجابة نبضية مؤلفة من إشارة النبضة الواحدة بالإضافة إلى نسخة عنها مزاحة وموزعة. يتكون التحويل التماثلي للانطواء من مرحلتين: في البداية يجب تحويل الانطواء إلى جداء وذلك باستخدام تحويل فورييه ويلي هذا مرحلة لوغاريتم يعيد عملية الجداء إلى عملية جمع عندئذ يتم فصل الإشارات بواسطة مرشح خطي ومن ثم يتم التراجع عن التحويل التماثلي.



الشكل (١-٢٠) منظومة تماثلية لفصل إشارات

نلاحظ في الشكل (١-٢١) تبديلا بالمواضع حيث إن المرشح الخطي يتعامل مع إشارات في مجال التردد بنفس الطريقة التي تتم فيها معالجة الإشارات في مجال الزمن، بعبارة أخرى يمكن القول إن الاستخدام الطبيعي لمجال الزمن ومجال التردد قد تبدل فمثلا إذا كانت مرحلة المرشح الخطي تستخدم الانطواء بواسطة FFT (توجد طيف الإشارات وتحسب جداؤها) فإن الطيوف التي حسب جداؤها هي عبارة عن إشارات في مجال الزمن. إن هذا التبادل بالأدوار قد أدى إلى ولادة مصطلحات جديدة وغريبة فعلى سبيل المثال يستخدم المصطلح (cepstrum) (وهو تغير في ترتيب أحرف كلمة spectrum كأن تقول فيض بدلا من طيف) للدلالة على تحويل فورييه للوغاريتم تحويل فورييه بنفس الشكل هناك مرشح تمرير طويل (long pass) ومرشح تمرير قصير (short pass) بدلا من مرشحات التمرير المرتفع والتمرير المنخفض.



الشكل (١-٢١)) منظومة تماثلية لفصل إشارات في المجال الترددي

يستخدم بعض المؤلفين مصطلح(frequency analysis) ومصطلح (filtering)، وهذا ليس إلا وصفا مبسطا لخوارزميات DSP المعقدة فالمعالجة التماثلية مليئة بالتفاصيل الدقيقة. فعلى سبيل المثال يجب أن يكون اللوغاريتم قادرا على التعامل مع إشارات الدخل الموجبة والسالبة وذلك لأن هذه إحدى ميزات إشارات الصوت يتطلب استخدام اللوغاريتم العقدي وهو مفهوم أكثر تطورا من اللوغاريتم العادي المستخدم في الحياة اليومية العلمية والهندسية؛ عندما يجبر المرشح الخطي أن يكون ذا صفة صفحة صفرية يتم إيجاد اللوغاريتم العقدي بأخذ اللوغاريتم العادي لإشارة الدخل بالقيمة المطلقة وبعد المرور في المرشح الخطي تعطى إشارة الخرج نفس الإشارة الجبرية للإشارة الأصلية. هناك مشكلة الانتحال التي تحدث عند حساب اللوغاريتم؛ تصور مثلا عملية تحويل إشارة جيبية إلى إشارة رقمية ووفقا لنظرية أخذ العينات يكفي أخذ عينتين أو أكثر من كل دور للتحويل، في الحقيقة أن لوغاريتم الإشارة الجيبية سيحوي تغيرات حادة (زوايا) وفي هذه الحالة نحتاج لعدد اكبر من العينات في الدور الواحد للمحافظة على شكل الوجه أي لتجنب الانتحال. قد يكون معدل التقطيع المطلوب اكبر بـ100 مرة من معدل التقطيع قبل اللوغاريتم والأكثر من ذلك إن تطبيق اللوغاريتم قبل التقطيع أو بعد التقطيع لن يغير في الأمر شيئا فالنتيجة هي نفسها: الانتحال سيحدث ما لم يكون معدل التقطيع كافيا لالتقاط التغيرات الحادة الناتجة عن عدم الخطية. النتيجة هي أن إشارة الصوت ستحتاج إلى معدل تقطيع حوالي 100 kHz بدلا من المعدل المعياري 8 kHz، حتى لو أنه تم حل هذه التفاصيل فليس هناك ضمان أن المرشح الخطي سيتمكن من فصل الإشارات التي تم تحويلها إلى إشارات خطية وذلك لأنه من الممكن أن تكون أطياف هذه الإشارات متداخلة حتى لو لم تكن أطياف الإشارات الأصلية متداخلة. تصور على سبيل المثال جمع إشارتين واحدة بتردد 1kHZ والأخرى بتردد 2kHz فطالما أن هذه الإشارة لا تتداخل في المجال الترددي سيكون بالإمكان فصلهما بشكل تام بواسطة مرشح خطي. تصور الآن أنه تم ضرب هاتين الإشارتين الجيبيتين في هذه الحالة سيؤدي استخدام معالجة تماثلية إلى أخذ لوغاريتم ناتج الجداء. ينتج عن هذا عملية أن جمع لوغاريتم الإشارة الجيبية سيكون له عدة توافقيات وبما إن توافقيات الإشارتين ستتداخل فإنه لن يكون بالإمكان فصلهما بشكل كامل [4].

الفصل الثاني

الدراسة النظرية لنظام التعرف على الصوت

٢-١ مقدمة

إن الهدف الأساسي من عملية تمييز الصوت تكون في تحديد المعلومات المتواجدة في الإشارة الصوتية، وبشكل مستقل عن التغيرات التي تنتج عن هوية الشخص الناطق، وأسلوب الكلام، والشروط المحيطة، وذلك لإيجاد الخوارزمية المناسبة لتوليد البصمة الصوتية لكل كلمة ومن ثم استخدامها في عملية التحكم. كما في حالة التحكم بكرسي للشخص المعاق عن طريق الصوت بحيث يستطيع الشخص المعاق التحكم بكافة حركات الكرسي عن طريق الأوامر الصوتية.

يوجد لدينا عدد من المصاعب لتوليد البصمة الصوتية وهي:

١- عملية النطق تختلف من شخص لأخر.

٢- تغيير النطق عند نفس الشخص.

٣- الضجيج الصوتي للبيئة المحيطة والتي تتداخل مع الإشارة الصوتية [1].

٢-٢ استخدامات التعرف الآلي على الكلام المنطوق

هناك عدة استخدامات في هذا المجال نذكر منها:

- مساعدة المكفوفين للتواصل مع الحاسب الآلي.
- الرد الرقمي على الاتصالات الواردة على الأنظمة الآلية.

٢-٣ التقنيات المستخدمة

إن أي نظام رقمي لتمييز الصوت لا بد من أن يمر بمراحل أساسية وهي:

١- كشف بداية ونهاية الكلمة.

٢- تخزين معطيات الكلمات الصوتية.

٣- استخلاص المعطيات الهامة من الإشارة الصوتية.

٤- تطبيق خوارزمية التطابق (قانون تمييز الصوت) [3].

٢-٤ كشف بداية ونهاية الكلمة

هناك عدة خوارزميات منها[5]:

٢-٤-١ نسبة التقاطع مع الصفر

يحسب نسبة التقاطع مع الصفر ZCR_t لإشارة الدخل من العلاقة التالية:

$$ZCR_t = \frac{1}{2}\sum_{n=1}^{N}\left| sign(x[n]) - sign(x[n-1]) \right| \qquad (٢-١)$$

حيث

$sign$: قطبية إشارة الدخل أي تكون قيمته ١ إذا كانت قطبية الإشارة موجبة و-١ إذا كانت قطبية الإشارة سالبة.

x : إشارة الدخل.

N:عدد عينات الإشارة.

فإذا كانت قيمة ZCR كبيرة فإن الإشارة تمثل الضجيج،أما إذا كانت قيمة ZCR صغيرة فإن الإشارة تمثل إشارة صوتية. تساعد هذه الخوارزمية في التمييز بين الإشارة والضجيج.

٢-٤-٢نسبة معدل عالي للعبور بالصفر

وهي تستخدم العلاقة التالية:

$$HZCRR = \frac{1}{2M} \sum_{t=0}^{M-1} \left| sign(ZCR_t - 1.5 \cdot avZCR) + 1 \right| \qquad (٢-٢)$$

ZCR_t :نسبة التقاطع مع الصفر في الإطار.

M:عدد الإطارات في الكلمة الصوتية.

$avZCR$:متوسط الـ ZCR في M إطار.

وهي تستخدم في مقاطع الموسيقى وذلك باعتبارها مكونة من كلمات متلاحقة [5].

٢-٤-٣ خوارزمية الصوت الأبيض

ترمز له بـ $v2wt$ وتعطى بالعلاقة التالية:

$$v2wt = 10\log \frac{\sum_{300Hz}^{4KHz} \left| X[k]^2 \right|}{\sum_{\forall k} \left| X[k]^2 \right|} \qquad (٢-٣)$$

X :مطال الطيف الترددي.

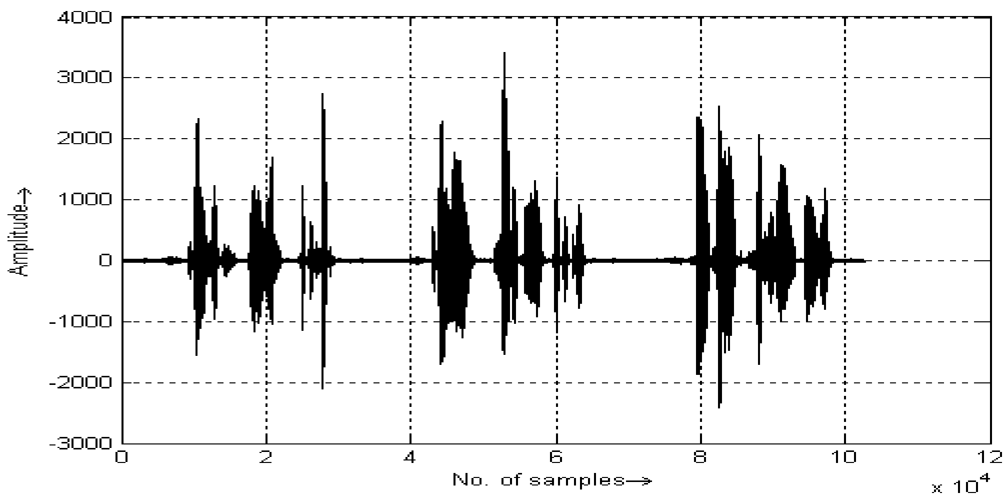
K:التردد.

وهي تقوم بحساب الطاقة المتواجدة في المجال (300-4000 Hz) إلى الطاقة الكلية في الإشارة. وهي ذات دقة عالية ولكنها تتطلب سرعة عالية للمعالج وذلك باعتباره يتطلب حساب الترددات ومن ثم تطبيق هذه الخوارزمية [5].

٢-٤-٤ تحديد بداية ونهاية الكلمة باستخدام الخواص الإحصائية للإشارة

٢-٤-٤-١ الخواص الإحصائية للإشارة الصوتية

إن الإشارة الصوتية تتغير ببطىء مع الزمن نسبيا أي عندما يتم فحص الإشارة الصوتية خلال فترات زمنية قصيرة (2 msec)فإن خصائصها مستقرة. ولكن خلال فترات أطول (200mSec) فإن هذه الخصائص تتغير نتيجة اختلاف الكلمات الصوتية وعادة 200msec الأولى من الإشارة الصوتية تعبر عن السكون (الضجيج) أي ما يقابل 1600عينة إذا كان معدل أخذ العينات 8000عينة في الثانية الشكل (٢-١)يوضح ذلك[3,8].



الشكل (٢-١) فترات الضجيج ضمن الإشارة الصوتية

تعتمد هذه الطريقة على استخدام المميزات الإحصائية للإشارة الصوتية. إن تابع التوزيع الاحتمالي يعطى بالعلاقة التالية:

$$P(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} \qquad (٢-٤)$$

حيث μ : المتوسط الحسابي.

x :العينة.

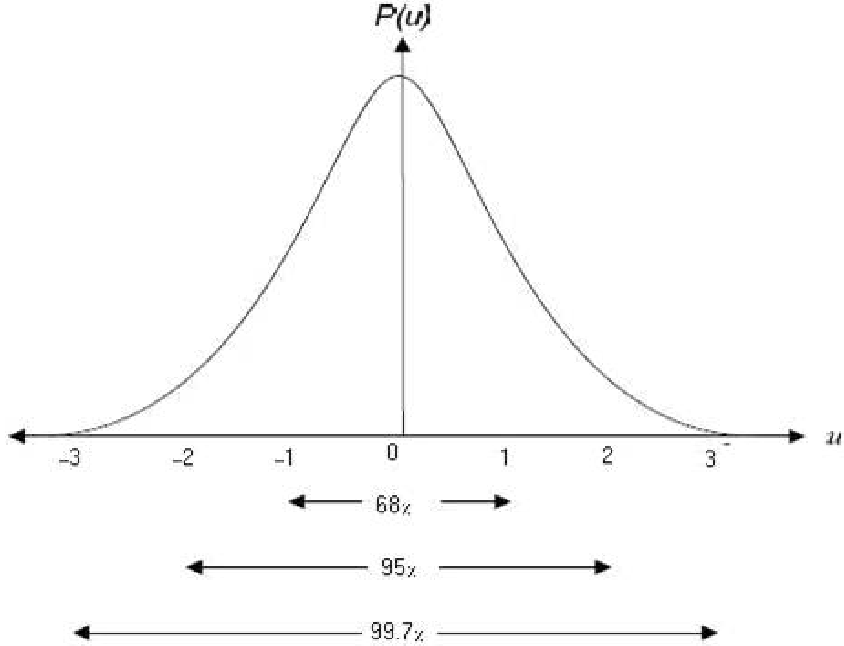
σ :الانحراف المعياري.

P :كثافة التوزع الاحتمالي.

كثافة تابع التوزع الاحتمالي تكون منحني على شكل جرس ،وهو يتحدد بقيم متحولين هما المتوسط الحسابي μ والانحراف المعياري σ^2 ،ويمكن كتابتها بالشكل $p(x) \approx N(\mu\sigma^2)$ ،كما إن المنحني يكون متناظر بالنسبة للمحور الشاقولي وبالنسبة لعينة ما x يمكن كتابة:

$$\left.\begin{aligned} &\Pr[|x-\mu| \leq \sigma] \triangleright 0.68 \\ &\Pr[|x-\mu| \leq 2\sigma] \triangleright 0.95 \\ &\Pr[|x-\mu| \leq 3\sigma] \triangleright 0.997 \end{aligned}\right\} \qquad (٢-٥)$$

والشكل (٢-٢) يوضح ذلك [6,7].



الشكل (٢-٢) منحني التوزيع الاحتمالي

حيث $u = x - \mu$

٢-٤-٤-٢ الخوارزمية المستخدمة في الطريقة الإحصائية

- يتم حساب المتوسط الحسابي لـ1600 عينة وفق العلاقة التالية :

$$\mu = \frac{1}{1600}\sum_{i=1}^{1600} x_{(i)} \qquad (٢-٦)$$

- الانحراف المعياري يعطى بالعلاقة التالية:

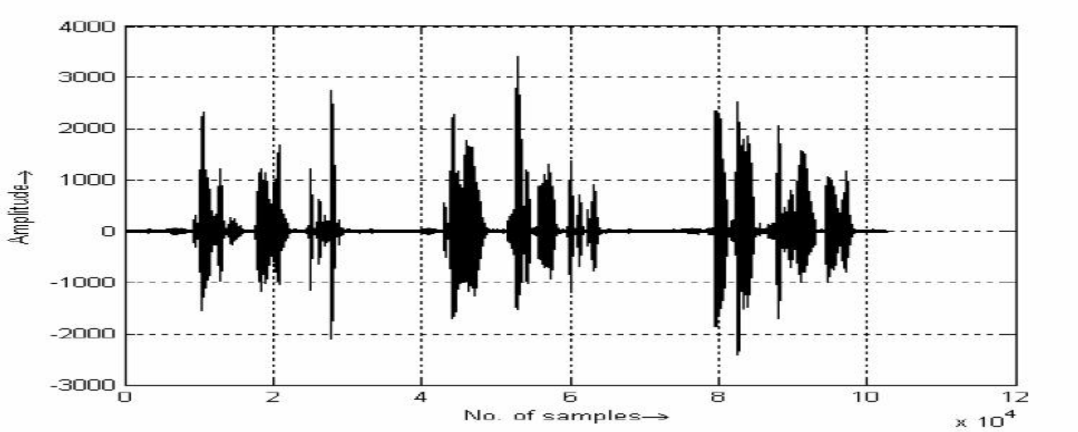
$$\sigma = \sqrt{\frac{1}{1600}\sum_{i=1}^{1600}(x_{(i)} - \mu)^2} \qquad (٢-٧)$$

- يتم اخذ العينة x ويطبق عليها العلاقة التالية :

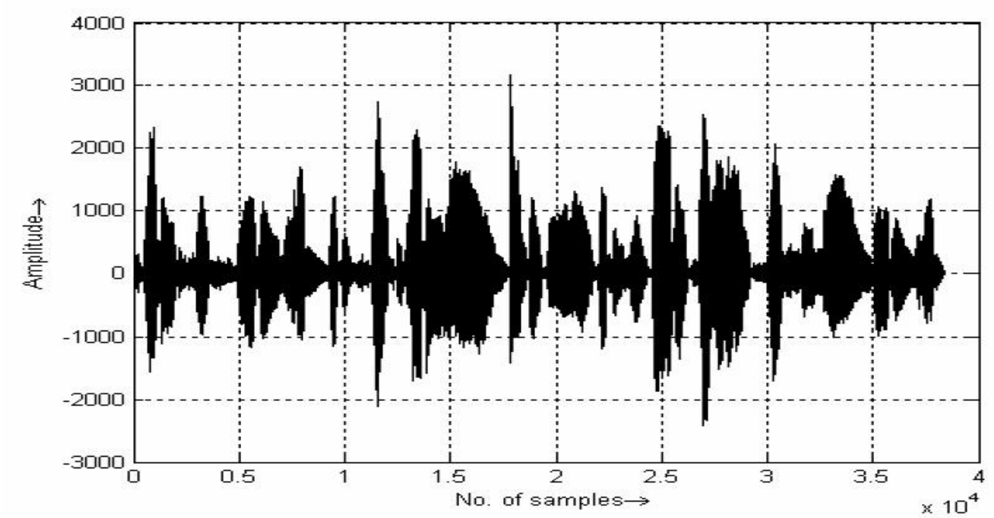
$$if\ \frac{|x-\mu|}{\sigma} > 3 \qquad (٢-٨)$$

- في حال أنه تحققت العلاقة السابقة يتم اعتبار العينة تمثل الإشارة الصوتية ويتم تخزينها وإلا فأنه يتم اعتبارها الضجيج.

الشكل(٢-٣) والشكل (٢-٤) يوضح كل من الإشارة قبل وبعد تحديد بداية ونهايتها.



الشكل (٢-٣) إشارة صوتية قبل تحديد بدايتها ونهايتها



الشكل (٢-٤) إشارة صوتية بعد تحديد بدايتها ونهايتها

يمكن تحديد مدى اقتراب العينات التي تم أخذها من العينات الأصلية بالعلاقة التالية :

$$\%correctness = 100 - \frac{N_{manual} - N_{a\lg orithm}}{N_{manual}} \times 100 \qquad (٢-٩)$$

حيث N_{manual}: عدد عينات الإشارة قبل تطبيق الخوارزمية عليها.

$N_{a\lg orithm}$: عدد عينات الإشارة بعد تطبيق الخوارزمية[10,11].

٢-٥تخزين الكلمات الصوتية (Storing Phonetic Words)

في هذه المرحلة سوف يتم تقسيم الكلمة الواحدة إلى مجموعة من الإطارات كل واحدة منها مكون من (500) عينة أي ما يعادل 50ms وذلك في حالة استخدام تردد أخذ العينات 10KHz.

وهناك طريقة أخرى وذلك في حالة استخدام سبعة مرشحات وذلك بتخزين قيمة الخرج لكل مرشح في ملف خاص به ولكن هذه الطريقة معقدة نوعا ما.

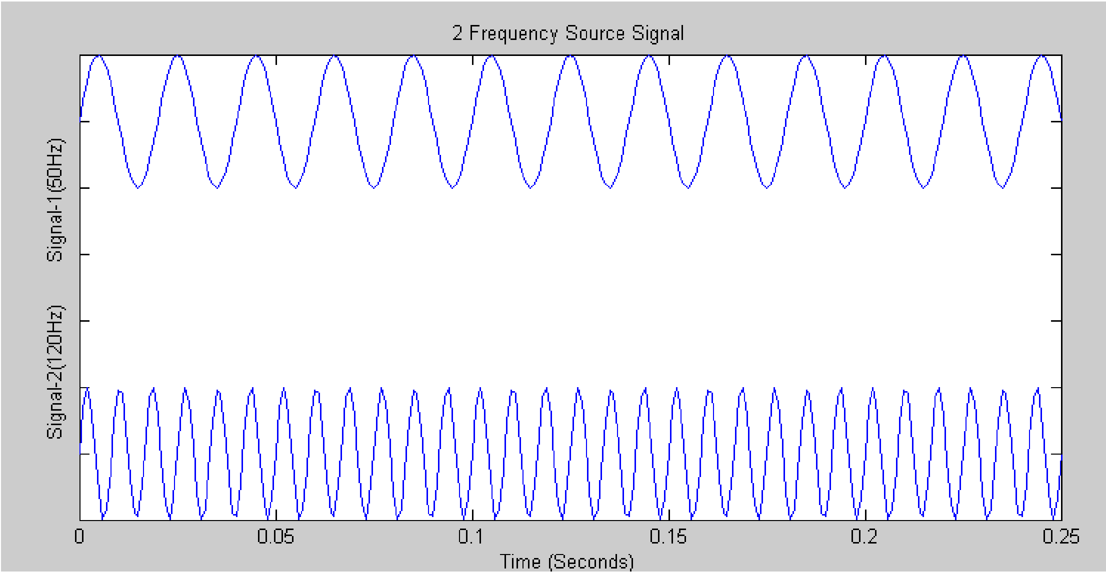
٢-٦استخلاص المعطيات الهامة من الإشارة الصوتية

من أهم الطرق المستخدمة في استخلاص المعطيات الهامة من الإشارة هي:

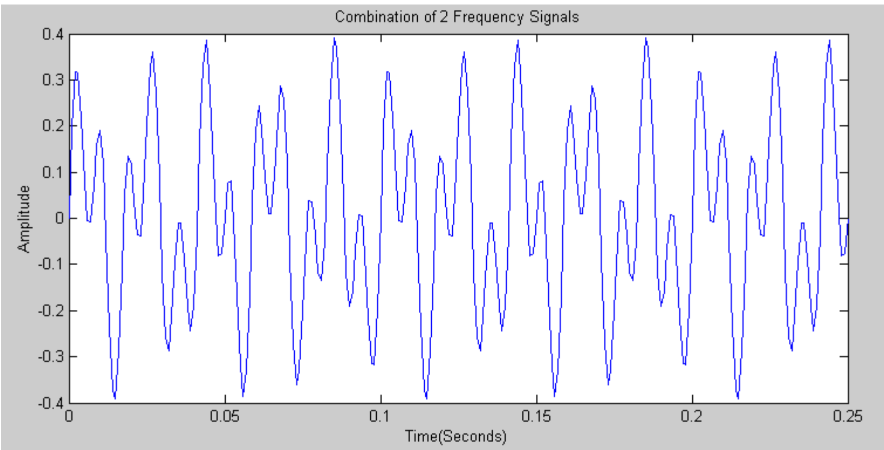
٢-٦-١تحليل فورية السريع FFT

هذه الطريقة تعتمد على تطبيق تحليل فورية السريع على الإشارة الصوتية وتحديد الترددات الأساسية فيها، ومن المعروف إن تحليل فورية يمثل عملية طوي للإشارة ولكن مع اختصار في عدد العينات لذلك لدى تنفيذ هذا التحويل على الكلمة الواحدة فإن زمن التنفيذ سوف يعتمد على سرعة المعالجة [9].

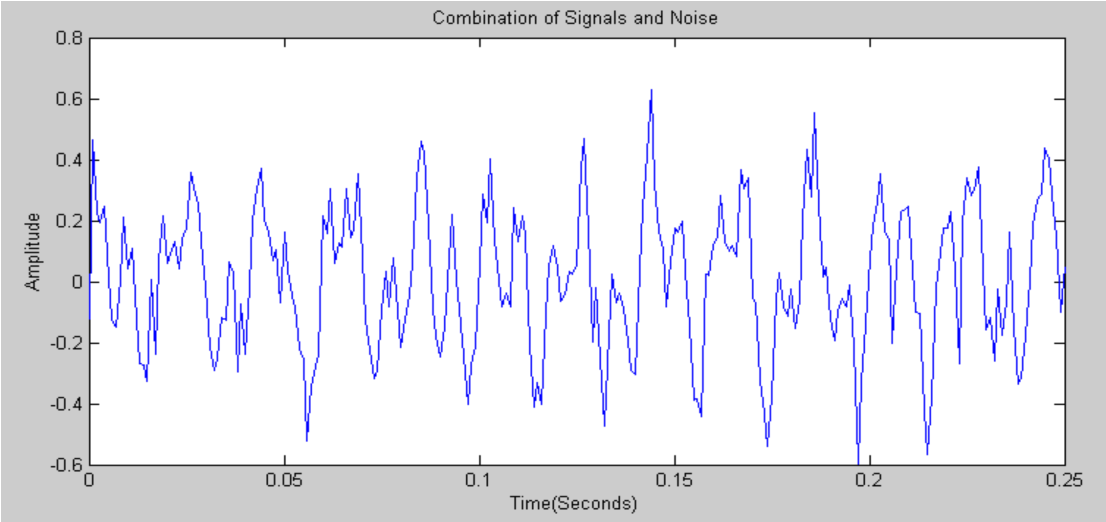
يبين الشكل (٢-٥)إشارة مكونة من إشارتين جيبيتن بترددين مختلفين 50Hz و 120Hz، في الشكل (٢-٦) يبين إشارة ضجيج،بإضافة الضجيج الناتج عن الوسط المحيط سوف ينتج إشارة صوتية كما في الشكل (٢-٧).



الشكل (٢-٥) إشارتين جيبيتين

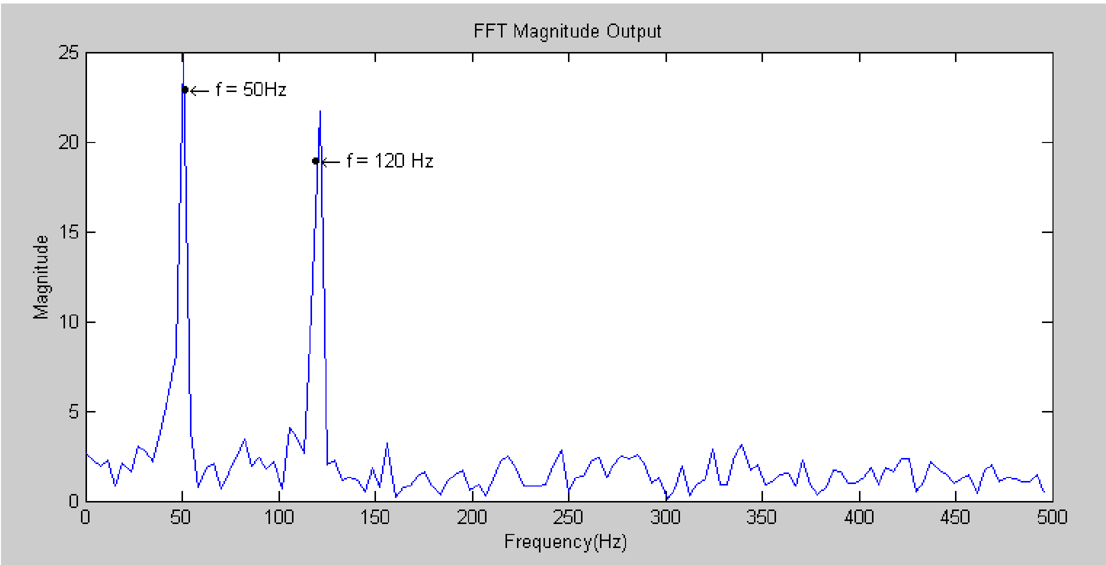


الشكل (٢-٦) يمثل إشارة صوتية ناتجة عن جمع الإشارتين في الشكل (٢-٥)



الشكل (٢-٧) الإشارة الصوتية في الشكل (٣-٦)بعد إضافة الضجيج إليها

بتطبيق تحويل فورييه السريع على الإشارة الصوتية سوف ينتج لدينا ترددين أساسيين هما 50 و 120 هرتز كما في الشكل (٢-٨) وبذلك يكفي أن يتم تخزين مطال هذين الترددين بدلا من تخزين الكلمة الصوتية بأكملها التي سوف تأخذ حيز كبير من الذاكرة، وبذلك سوف يتم اختصار الزمن اللازم لإجراء عملية التطابق لدى تطبيق أحد خوارزميات التطابق.



الشكل (٢-٨) الطيف الترددي للإشارة الصوتية في الشكل (٢-٧)

٢-٧ خوارزمية تمييز الصوت

٢-٧-١خوارزمية LMS (Least Mean Square)

لنفرض أنه لدينا العلاقة بين x و y ممثلة بنقاط؛ إذا أردنا إن نرسم المستقيم التقريبي لهذه النقاط سوف نحصل على حلول مختلفة، ولكن هناك طريقة رياضية لإيجاد أفضل مستقيم تقريبي لهذه النقاط تدعى LMS. مبدأ هذه الطريقة يعتمد على إيجاد خط مستقيم والذي يعطي اصغر قيمة لمربعات الانحرافات بين النقاط لهذا التابع وهذا المستقيم التقريبي.

يجب أن يكون الانحراف أصغري

$$\sum d_i^2 = Min \qquad (٢-١٠)$$

$$d_1 = (d_1 y - d_1^{'} y)^2 \qquad (٢-١١)$$

حيث d:بعد النقاط المطلوب أيجاد معادلتها عن نقاط معادلة المستقيم.

في حالة دخل واحد وخرج واحد نعتبر المستقيم الأمثلي هو:

$$y = ax + b \qquad (٢-١٢)$$

والمطلوب هو إيجاد قيم a وb الأمثلة ولإيجاد a وb نعتمد طريقة LMS

$$\sum e^2 = \min imum \qquad (٢-١٣)$$

$$\sum_{i=1}^{n} (ax_i + b - y_i)^2 = \min$$

$$\frac{dE}{db} = 0$$

$$\frac{dE}{da} = 0$$

$$b = \frac{(y_1 + y_2 + ... + y_n) - a(x_1 + x_2 + ... + x_n)}{n} \qquad (٢-١٤)$$

$$(ax_1 + b - y_1) + (ax_2 + b - y_2) + ... + nb = 0$$

$$b = \overline{y} - a\overline{x}$$

$$\frac{dE}{da} = 2\sum (x_i - \overline{x})(a(x_i - \overline{x}) + \overline{y} + y_i)$$

$$a = \frac{\sum (y_i - \overline{y})(x_i - \overline{x})}{\sum (x_i - \overline{x})^2} \qquad (٢-١٥)$$

$$b = \overline{y} - a\overline{x}$$

يعطى عامل الارتباط بالعلاقة التالية:

$$R = \frac{\sum (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum (x_i - \overline{x})^2 \sum (y_i - \overline{y})^2}} \qquad (٢-١٦)$$

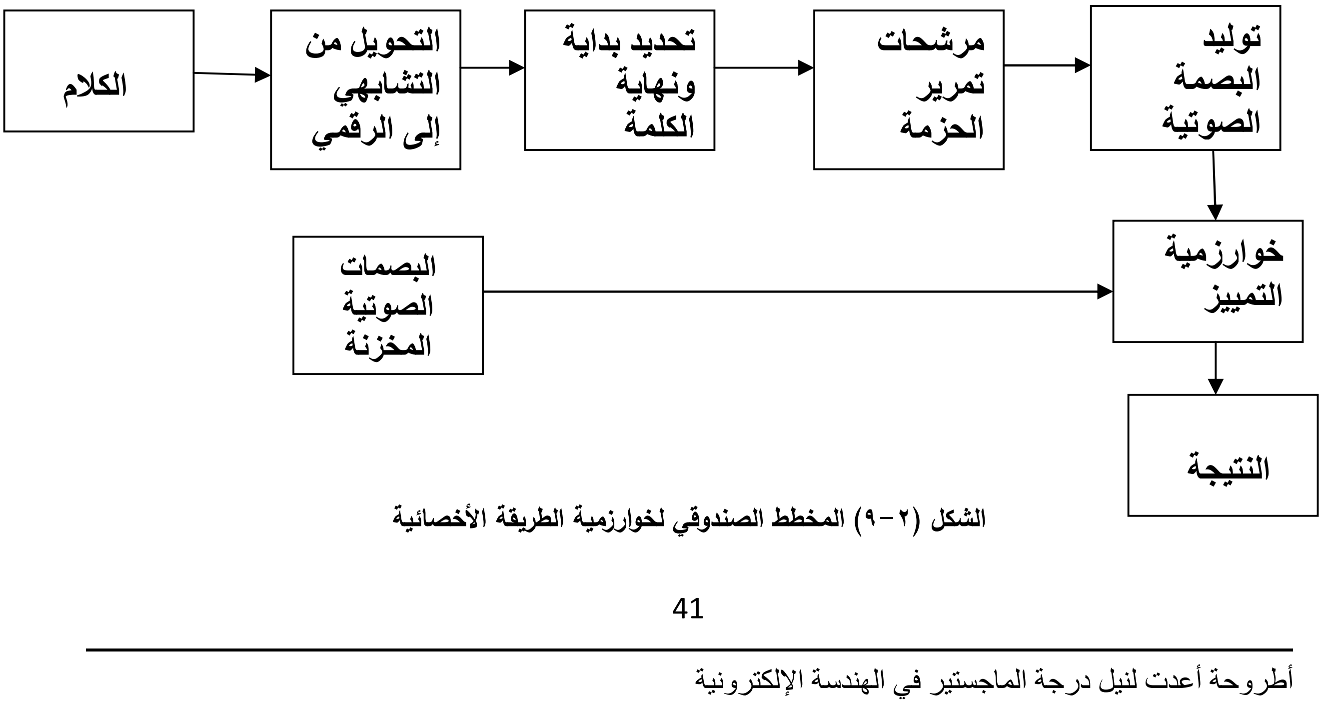
حيث $\overline{y}$:المتوسط الحسابي للعينات y.

$\overline{x}$:المتوسط الحسابي للعينات X .

حيث إن قيمة هذا العامل تتراوح بين +١ و-١ وكلما كانت القيمة المطلقة لهذا العامل تتناهى إلى الواحد كانت دقة التقريب أفضل.يمكن تطبيق هذه الخوارزمية لإيجاد مقدار التطابق بين البصمات الصوتية للكلمات المخزنة والكلمات المنطوقة [6].

٢-٨ الطريقة الإحصائية

يمكن الآن إن نوضح المخطط الأساسي لهذه الخوارزمية في الشكل (٢-٩)

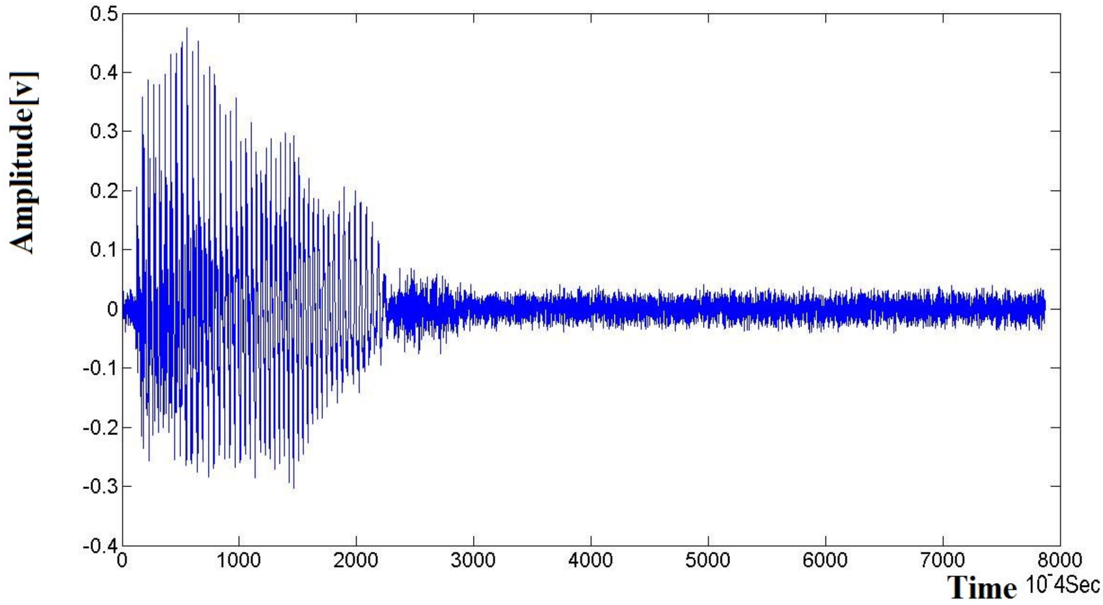


الشكل (٢-٩) المخطط الصندوقي لخوارزمية الطريقة الأخصائية

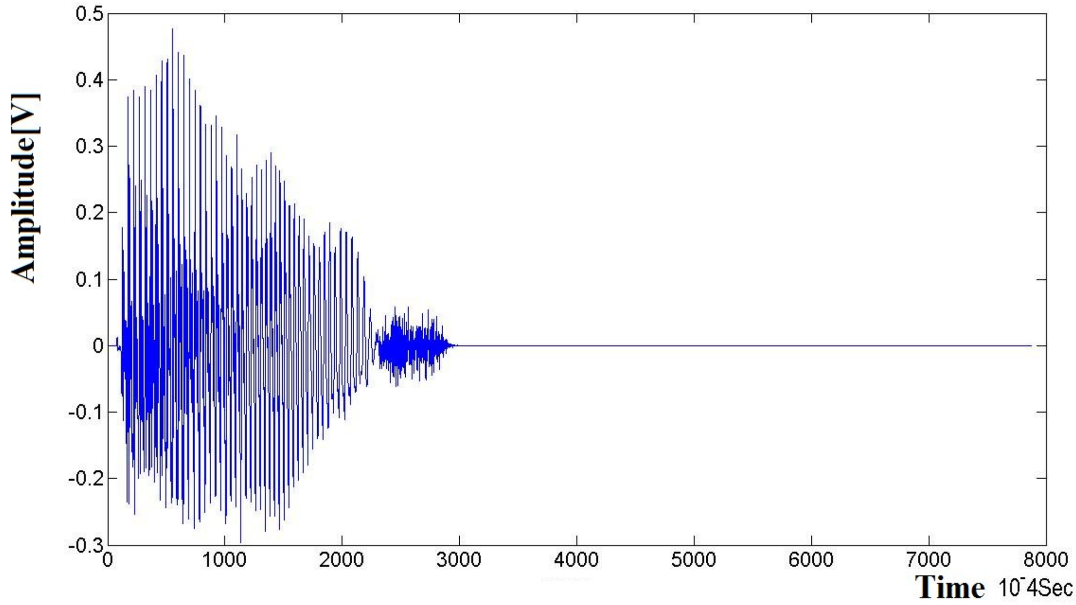
٢-٨-١تحديد بداية ونهاية الكلمة

في البداية لابد من تحديد بداية ونهاية الكلمة الصوتية، لتحقيق ذلك استخدمنا الطريقة الإحصائية لأنها تعطي نتائج جيدة لفصل الإشارة عن الضجيج ولا تستغرق زمنا طويلا.

الشكل(٢-١٠) يوضح انه تم تحديد بداية ونهاية الكلمة حيث نلاحظ في الشكل (٢-١٠-a) يسبق بداية الكلمة إشارة الضجيج ولكن في الشكل (٢-١٠-b) تم حذف هذا الضجيج وتم تحديد بداية الكلمة .



الشكل (٢-١٠-a) الإشارة الصوتية قبل حذف الضجيج



الشكل(٢-١٠-b) الإشارة الصوتية بعد حذف الضجيج

٢-٨-٢الترشيح الرقمي للإشارة

لتحليل الصوت نحتاج لإيجاد الترددات الموجودة في الكلمة المكتشفة، ولتحقيق ذلك سوف نستخدم مرشحات تمرير الحزمة من المرتبة الرابعة من نوع تشيبيشف (Chebyshev). ولبناء مرشح من المرتبة الرابعة يجب وصل مرشحين على التتالي من المرتبة الثانية للحصول على دقة جيدة وذلك باستخدام المعادلات التالية:

$$y_1(n)=b_{11}x_1(n)+b_{12}x_1(n-1)+b_{13}x_1(n-2)-a_{11}y_1(n-1)-a_{12}y_1(n-2) \quad (٢-١٧)$$

$$y_2(n)=b_{21}x_2(n)+b_{22}x_2(n-1)+b_{23}x_2(n-2)-a_{21}y_2(n-1)-a_{22}y_2(n-2) \quad (٢-١٨)$$

$$y_{out}(n) = gy_2(n) \quad (٢-١٩)$$

حيث إن المعاملات a,b,g تم حسابها باستخدام برنامج Matlab [4].

٢-٨-٣توليد البصمة الصوتية

في البداية نفترض إن الكلمة المأخوذة من المدخل التشابهي قد تم تحديد بدايتها ونهايتها . يتم تمرير الإشارة بعد ذلك عبر مرشحات تمرير حزمة، ويتم تخزين خرج كل مرشح بشكل تراكمي ضمن البصمة الصوتية، التي تتكون من خرج المرشحات الرقمية، أي يمكن تخزين خرج كل مرشح في ست عشرة نقطة فقط. بذلك يتكون لدينا حوالي 160 نقطة معطيات لكل كلمة، و كل نقطة معطيات تتكون من 125 عينة، في حال استخدامنا لعشر مرشحات، بهذه الطريقة نكون قد وفرنا حجماً كبيراً من الذاكرة و دون استغراق وقت طويل.

٢-٨-٤حساب مقدار التطابق بين الكلمات الصوتية

لتحديد مقدار التطابق بين البصمة الصوتية للكلمة المكشوفة والبصمة الصوتية للكلمات المخزنة نقوم بتطبيق خوارزمية LMS من خلال علاقة المعامل R لإيجاد البعد الأصغري لكل نقطة من الكلمة المراد مقارنته مع الكلمة المخزنة يعطى المعامل R بالعلاقة التالية:

$$R=\frac{\sum (x_i-\overline{x})(y_i-\overline{y})}{\sqrt{\sum (x_i-\overline{x})^2 \sum (y_i-\overline{y})^2}} \quad (٢-٢٠)$$

حيث x :عينات البصمة الصوتية المخزنة.

y :عينات البصمة الصوتية للكلمة التي يتم النطق بها.

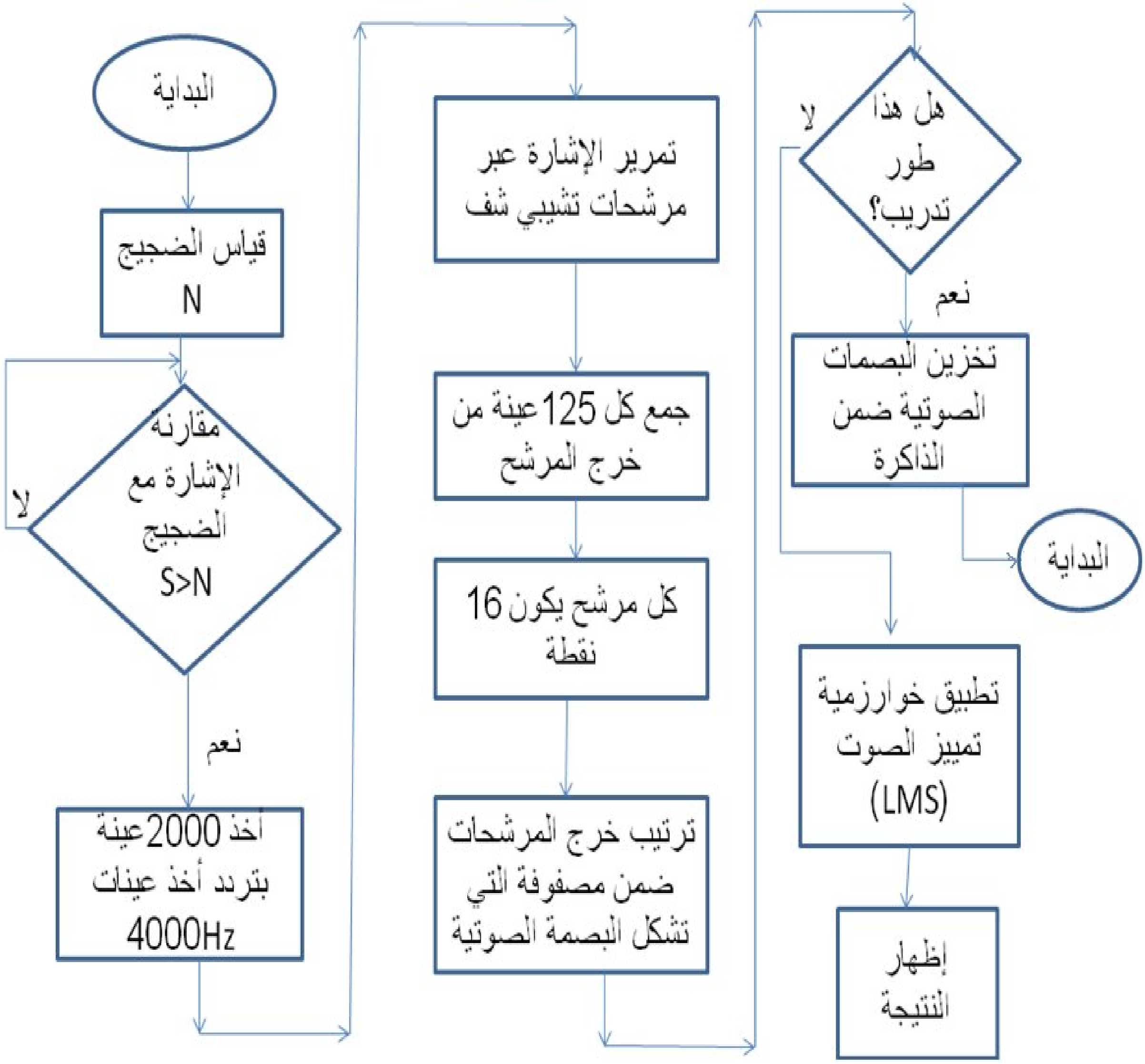
إن النقطة الأساسية في هذه الخوارزمية تكون في فحص المبدل ADC الذي يمثل مدخل الإشارة الصوتية بتردد 4KHZ إذا كانت قيمة المبدل اكبر من قيمة العتبة يتم البدء بتخزين الكلمة ذات الطول 0.5 ثانية بالتالي فان الكلمة الواحدة تحتاج إلى 2000 عينة فقط. بعدها يتم تمرير الكلمة عبر عشر مرشحات تمرير الحزمة ثم يتم تخزينها ضمن البصمة الصوتية بهذه الطريقة يتم تخزين البصمة الصوتية لكل كلمة ليتم مقارنتها لاحقا مع الكلمة المكشوفة بالتالي الاختلاف الأقل يمثل التطابق بين الكلمتين [6].

الفصل الثالث

المحاكاة باستخدام برنامج Matlab

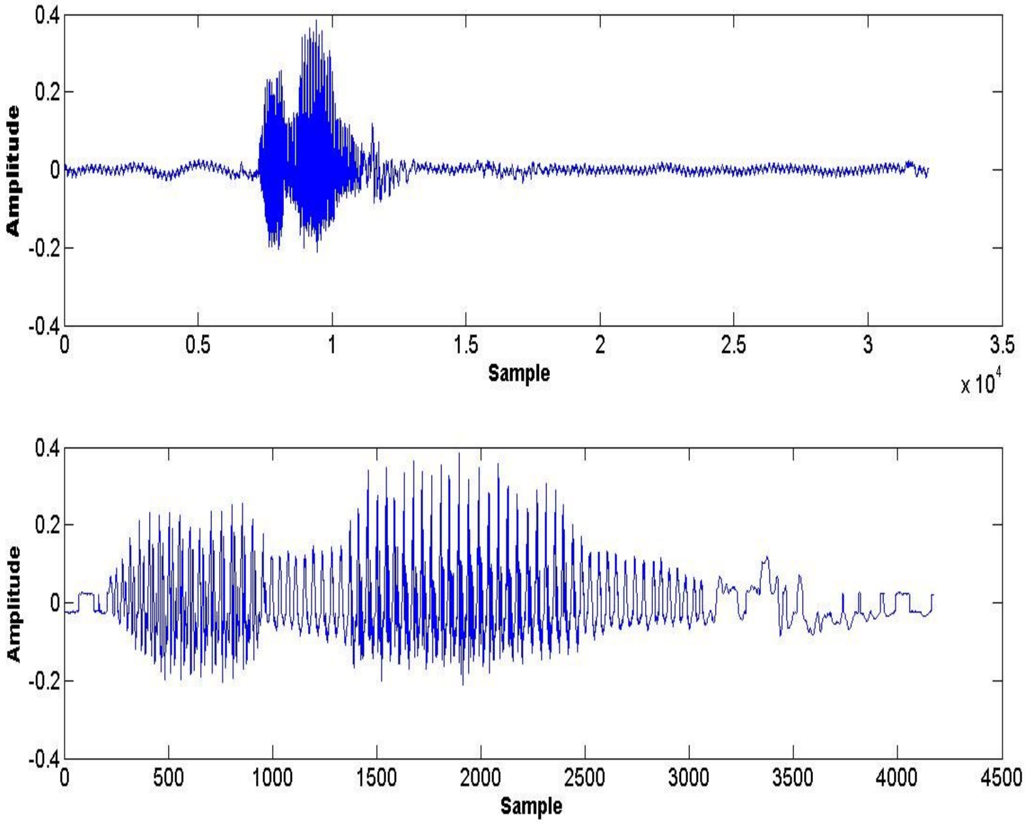
النتائج والمناقشة

للحصول على نتائج الدراسة السابقة قمنا بتنفيذ هذه الخوارزمية باستخدام برنامج Matlab لأجراء عملية المحاكاة، والشكل (٣-١) يبين المخطط الصندوقي لبرنامج المحاكاة



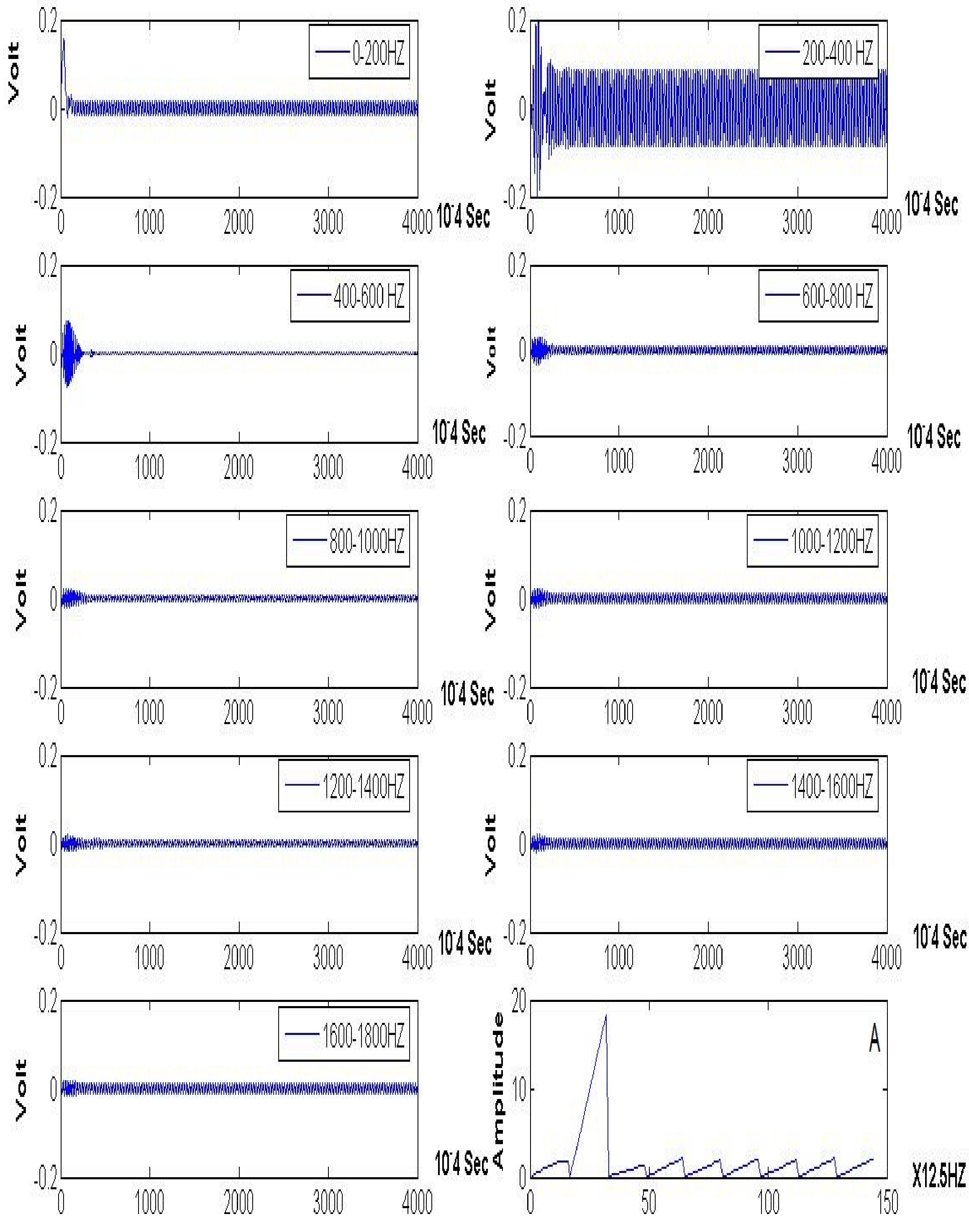
الشكل (٣-١) المخطط الصندوقي لبرنامج المحاكاة

الشكل (٣-٢) يوضح تحديد بداية ونهاية الكلمة حيث نلاحظ انه تم حذف الضجيج. قبل تطبيق الخوارزمية عدد العينات حوالي 32500 عينة وبعد تطبيق الخوارزمية أصبحت 4000 عينة.



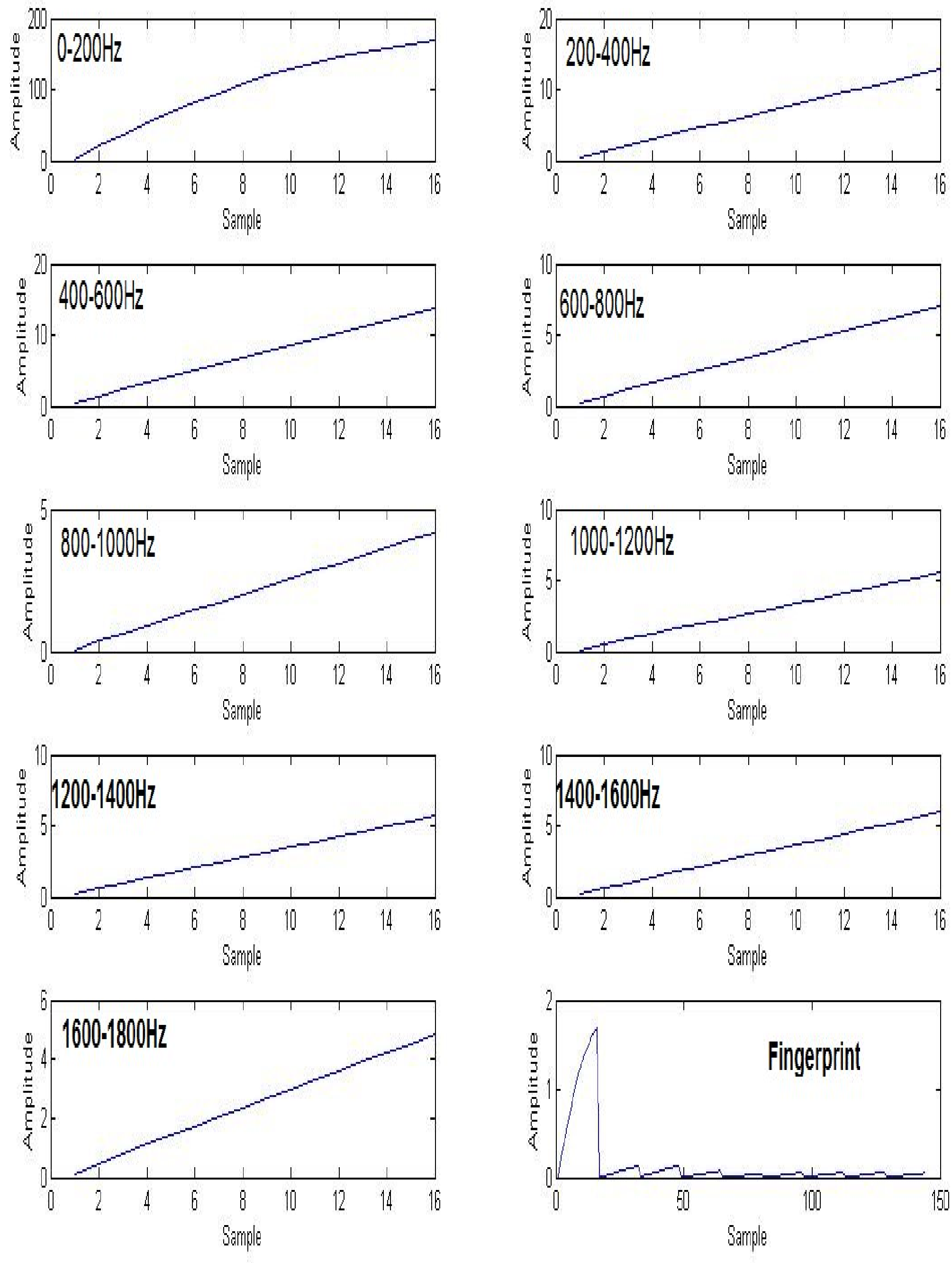
الشكل (٣-٢) عملية تحديد بداية ونهاية الكلمة

الشكل (٣-٣) يوضح عملية الترشيح الرقمي للإشارة الصوتية،حيث تمرر الإشارة عبر تسع مرشحات رقمية من نوع تشيب شف، المجال الأول من 0-200 Hz والثاني180-400Hz وهكذا حتى المرشح التاسع من 1580-1800 Hz، وأخيرا المرشح العاشر هو مرشح تمرير عالي بتردد قطع 1800 Hz.



الشكل (٣-٣) يوضح عملية الترشيح الرقمي للإشارة الصوتية بتطبيق الخوارزمية الموضوعة ضمن مجالات ترددية مختلفة؛ والمنحني A يوضح البصمة الصوتية

الشكل (٣-٤) يوضح خرج المرشحات الرقمية التسع والبصمة الصوتية. بما أن الإشارة الصوتية المأخوذة تحتوي على 4000 عينة فإن خرج كل مرشح يتكون من 4000 عينة. أيضا يتم تخزين خرج كل مرشح في 16 نقطة، أي يتم جمع 125عينة في خرج المرشح بشكل تراكمي في نقطة واحدة؛ هذا يعني إن كل 12.5 Hz تقابل نقطة واحدة على محور السينات .

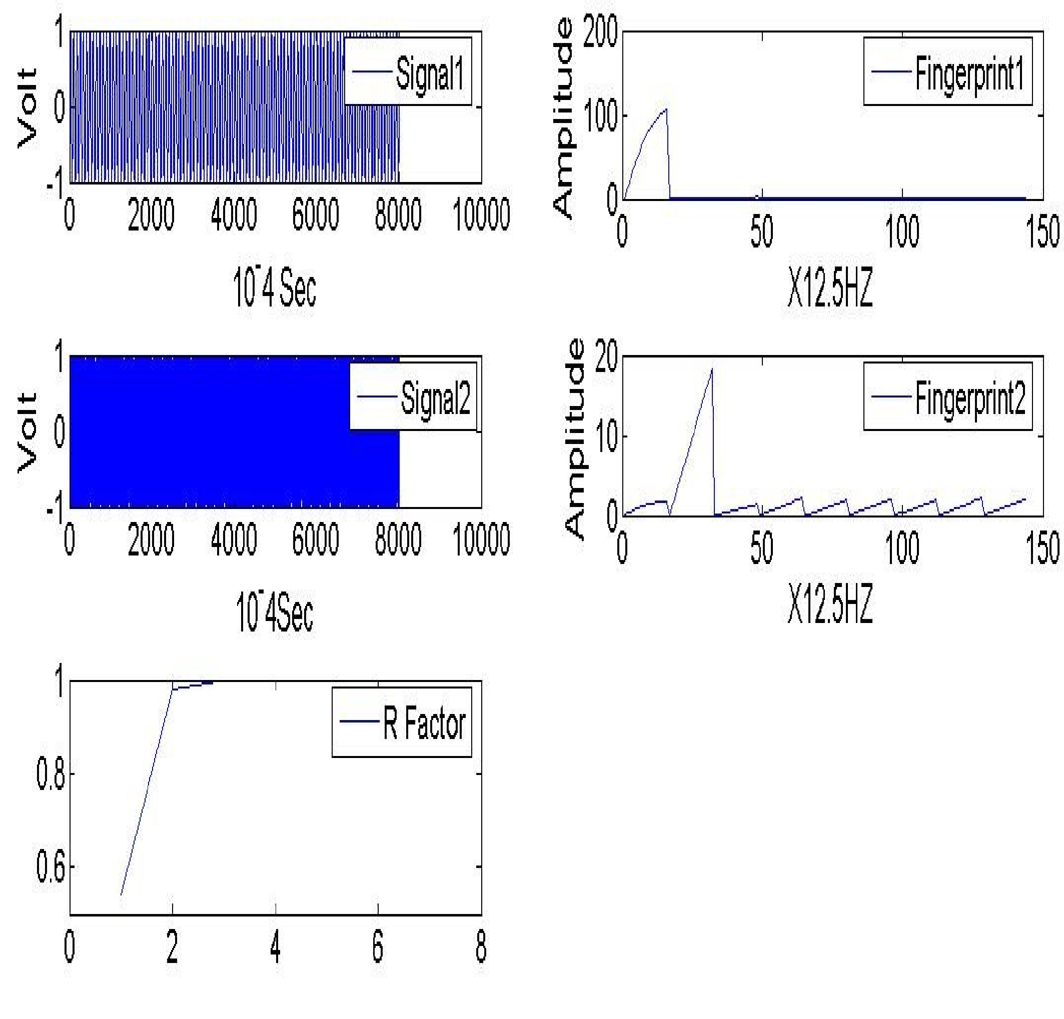


الشكل (٣-٤) يمثل خرج المرشحات الرقمية ضمن مجالات ترددية مختلفة ،والبصمة الصوتية (Fingerprint)

لإيجاد مقدار التطابق بين الإشارتين -Signal-1 و -Signal-2 اللتان تمثلان إشارتين جيبيتين بتردد 100 HZ و300 HZ على التتالي.

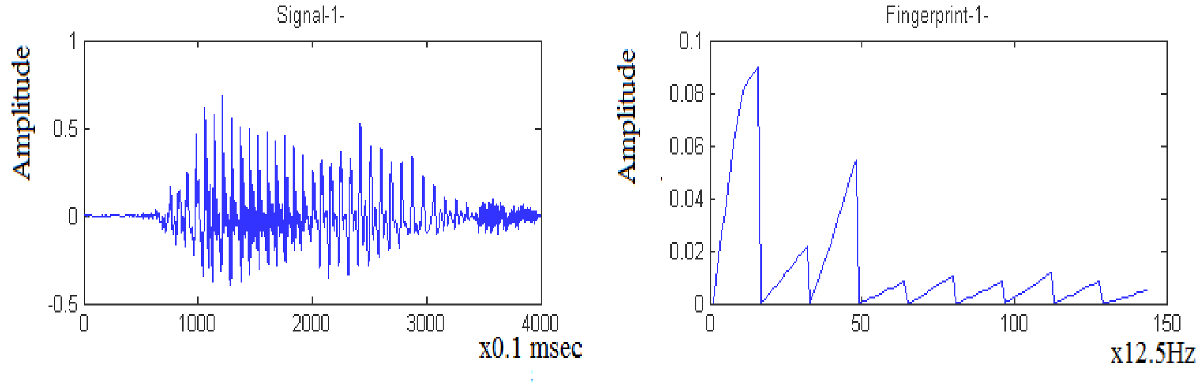
و-Fingerprint-1 &-Fingerprint-2 يمثلان البصمة الصوتية للإشارتين على التتالي حيث نلاحظ انه لكل منهما تردد يختلف عن الآخر.

والمنحني R يظهر عدم التطابق بين هاتين الإشارتين لأنه في حالة التطابق المقدار R يأخذ قيمة (١) في الشكل (٣-٥) يوضح ذلك .



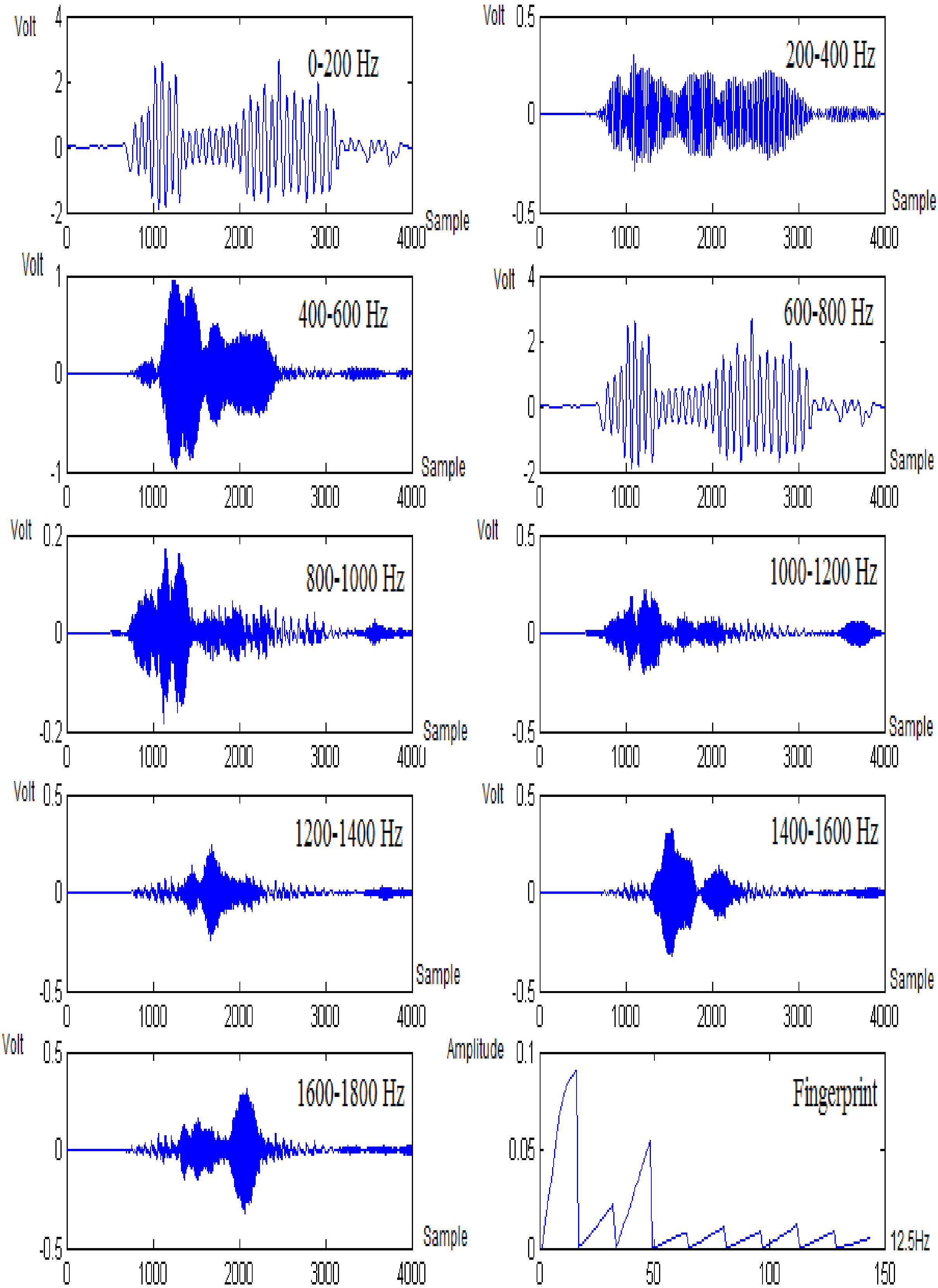
الشكل(٣-٥) إشارتين جيبيتن مع البصمة الصوتية وقيمة المعامل Rالذي يوضح عدم التطابق بين البصمتين الصوتيتين لأن قيمته لا تأخذ الواحد من أجل كل العينات

أما الشكل (٣-٦) يمثل الإشارة الصوتية للكلمة Orange وكذلك الطيف الترددي لها حيث نلاحظ أن الطيف الترددي لها يختلف عن السابق لأنه يحتوي على عدد من الترددات.



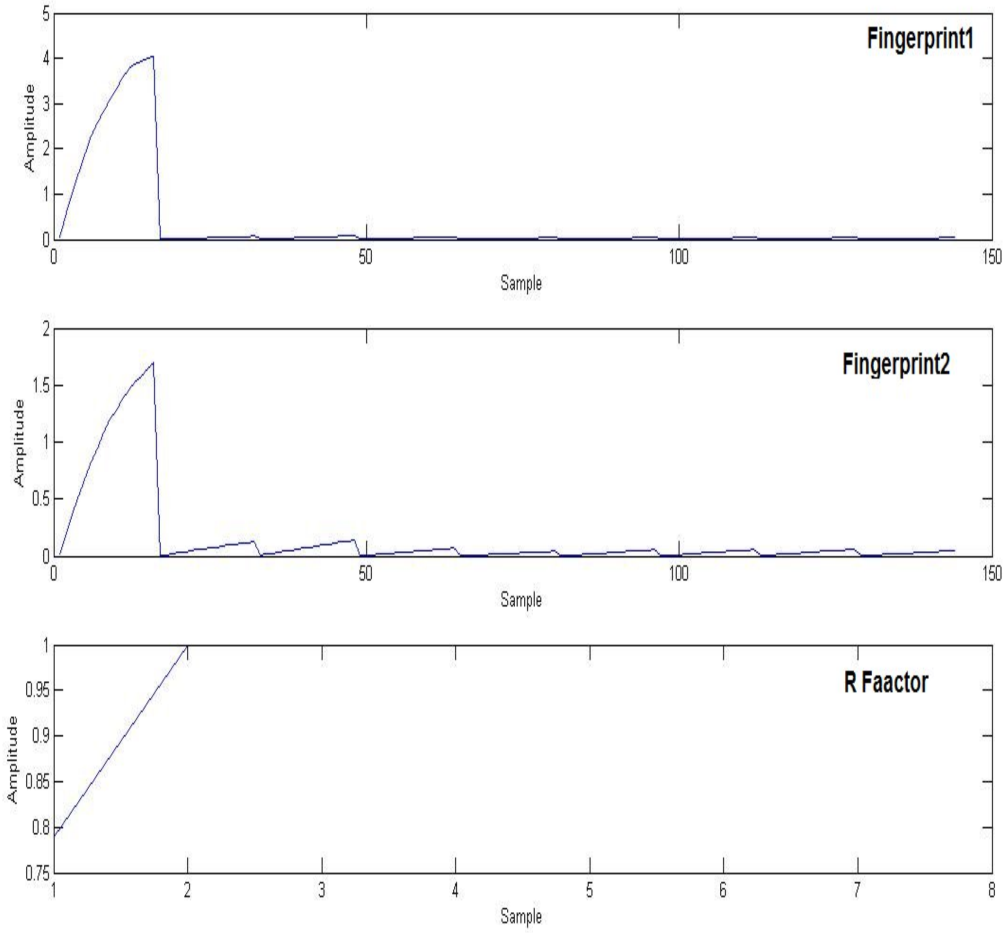
الشكل (٣-٦) الكلمة الصوتية مع البصمة الصوتية لكلمة Orange

الشكل (٣-٧) يوضح الطيف الترددي لكلمة Orange وخرج المرشحات الرقمية نلاحظ إن الطيف الترددي يتركز في المجال 0-200 Hz و580-800 Hz و 1180-1400 Hz هرتز كما إن المجالات فيما بينها أيضا مختلفة وهذا الاختلاف واضح في البصمة الصوتية.



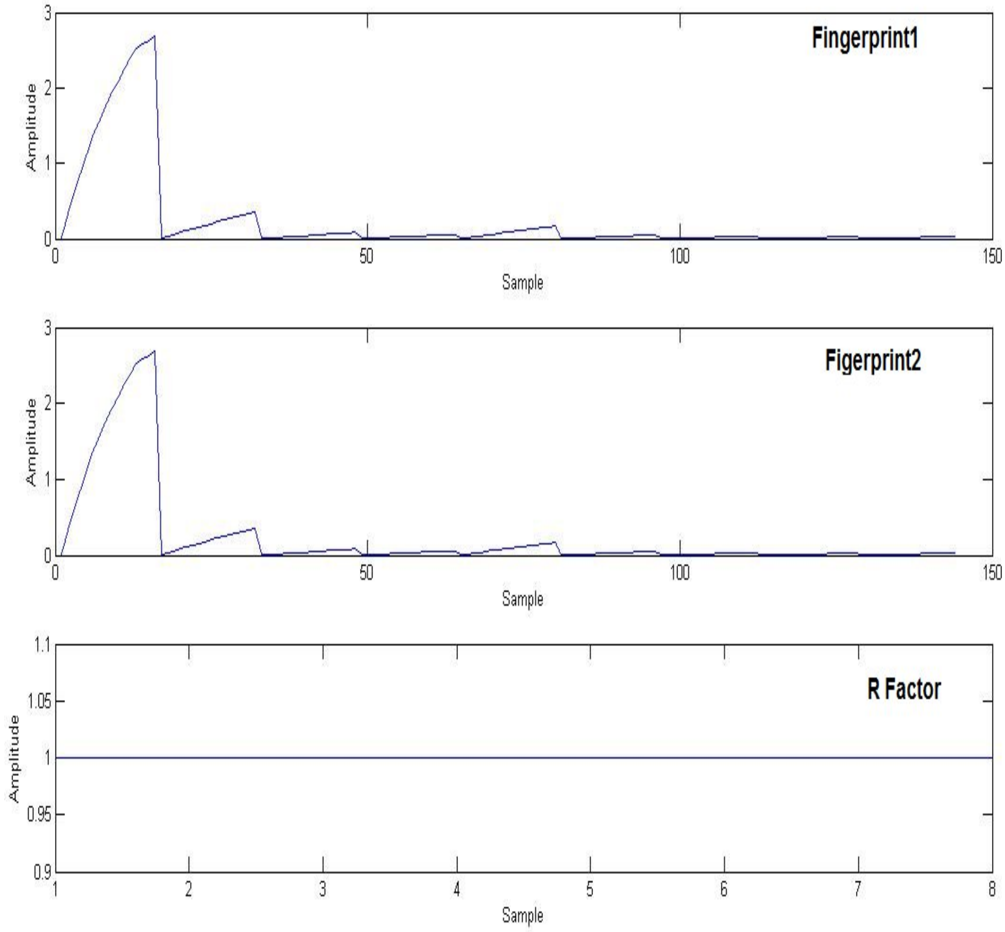
الشكل (٣-٧) الطيف الترددي لكلمة Orange ضمن مجالات ترددية مختلفة مع البصمة الصوتية(Fingerprint)

في الشكل (٣−٨) نلاحظ قيمة المعامل R باستخدام الخوارزمية LMS وهو يشير إلى عدم التطابق بين الإشارتين لأن قيمة المعامل R لا تساوي الواحد.



الشكل (٣−٨) يوضح قيمة المعامل R الذي يعطي عدم التطابق بين البصمة الصوتية الأولى والثانية

أما الشكل (٣−٩) فإن قيمة المعامل R يشير إلى التطابق التام بين الإشارتين حيث أن قيمة المعامل تساوي الواحد.



الشكل (٣-٩) يوضح التطابق بين البصمة الصوتية الأولى والثانية، قيمة المعامل R تساوي الواحد

الفصل الرابع

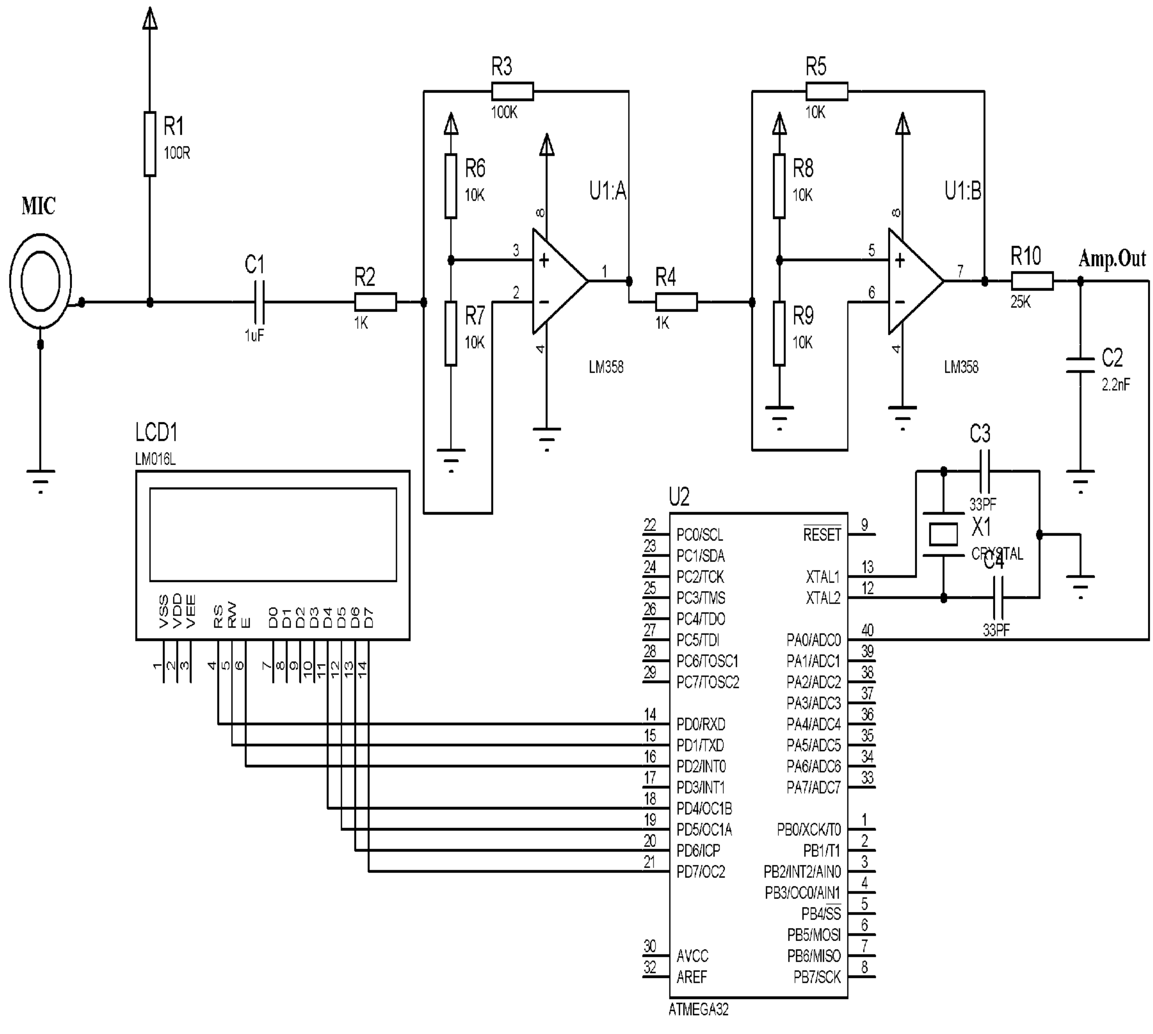
الدارة العملية لتمييز الصوت

الدارة العملية

يوضح الشكل (٤–١)الدارة التي تم تنفيذها وهي تتكون من الأجزاء الأساسية التالية:

١. الميكروكونترولر MEGA32 .

٢. دارة تضخيم الصوت والتي تتكون من المضخم العملياتي LM358 .



الشكل (٤–١) يبين دارة تمييز الصوت باستخدام الميكروكونترولر

من خلال هذه الدارة تم تنفيذ الخوارزمية الإحصائية التي توصلنا إليها في الفصل الثاني وتم برمجة MEGA32 وفق برنامج المحاكاة في الفصل السابق؛تم تجريب عدد من الكلمات ضمن غرفة كانت قيمة الضجيج فيها صغيرة 40dB- جداً وقد تم حسابها من خرج دارة التضخيم بواسطة راسم إشارة.وقد كانت النتائج كما هو مبين في الجدول (٤–١).

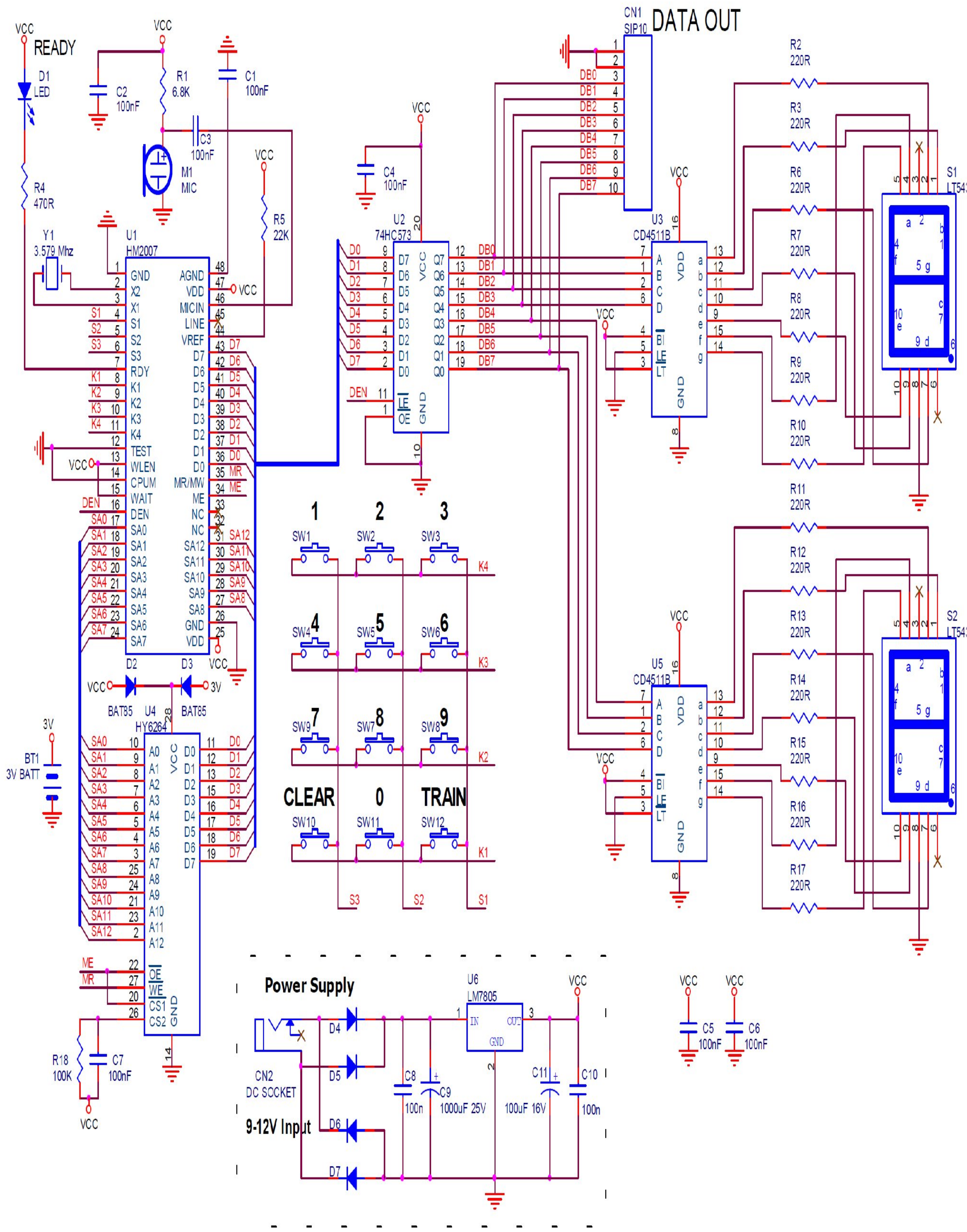
الجدول (٤-١)

الكلمة	عدد مرات التجريب	نسبة التمييز للشخص الأول	نسبة التمييز للشخص الثاني
Hello	20	23%	22%
Left	20	28%	30%
Right	20	23%	23%
Orange	20	29%	29%
Me	20	38%	35%
Car	20	33%	31%
Bare	20	34%	32%
Salad	20	28%	27%
Gat	20	40%	41%
Bid	20	40%	42%
أمام	20	32%	31%
يمين	20	30%	30%
يسار	20	31%	32%
ثانية	20	24%	23%
سبب	20	35%	34%
قال	20	41%	39%
جال	20	41%	38%

إن المتوسط الحسابي الكلي لنسبة التمييز الصوت 32% وتعتبر هذه النتيجة قليلة؛ علماً إن بعض الكلمات التي تحوي على حرف انفجاري أعطت نسبة تمييز اكبر، ولكن مع ذلك تعتبر قليلة ويعود السبب في إن الميكروكونترولر يملك إمكانيات محدودة في مجال الإشارة التشابهية؛ لذلك سوف نقوم باستخدام دارة متكاملة تقوم بعملية تمييز الصوت وهذا ما ينصح به بعض المراجع العلمية [12] . لأن النظام سوف يتم استخدامه لمساعدة إنسان معاق لذلك يجب أن يتمتع بدقة جيدة.

الشكل (٤-٢) يوضح دارة تمييز الصوت، هذه الدارة تعتمد بشكل أساسي على الدارة المتكاملة HM2007. يمكن لهذه الشريحة تمييز عشرين كلمة صوتية على الأقل بعد أن يتم تلقينها للنظام،أقصى مدة زمنية للكلمة الصوتية 1.9Sec.

القسم الثاني من الدارة تتمثل بالميكروكونترولر MEGA32،الذي يستخدم خرج الدارة HM2007 لإخراج الأوامر إلى الأجهزة المحيطة.



الشكل (٤-٢) دارة تمييز الصوت

٤-١مكونات الدارة

تتكون الدارة من:

١. الشريحة HM2007.
٢. ذاكرة RAM بحجم 8KB.
٣. لوحة مفاتيح من 0 إلى 9 ومفتاحي Train و Clear.
٤. شاشة إظهار تتألف من خانتين.

٤-٢ الشريحة HM2007

خصائص هذه الشريحة هي كما يلي:

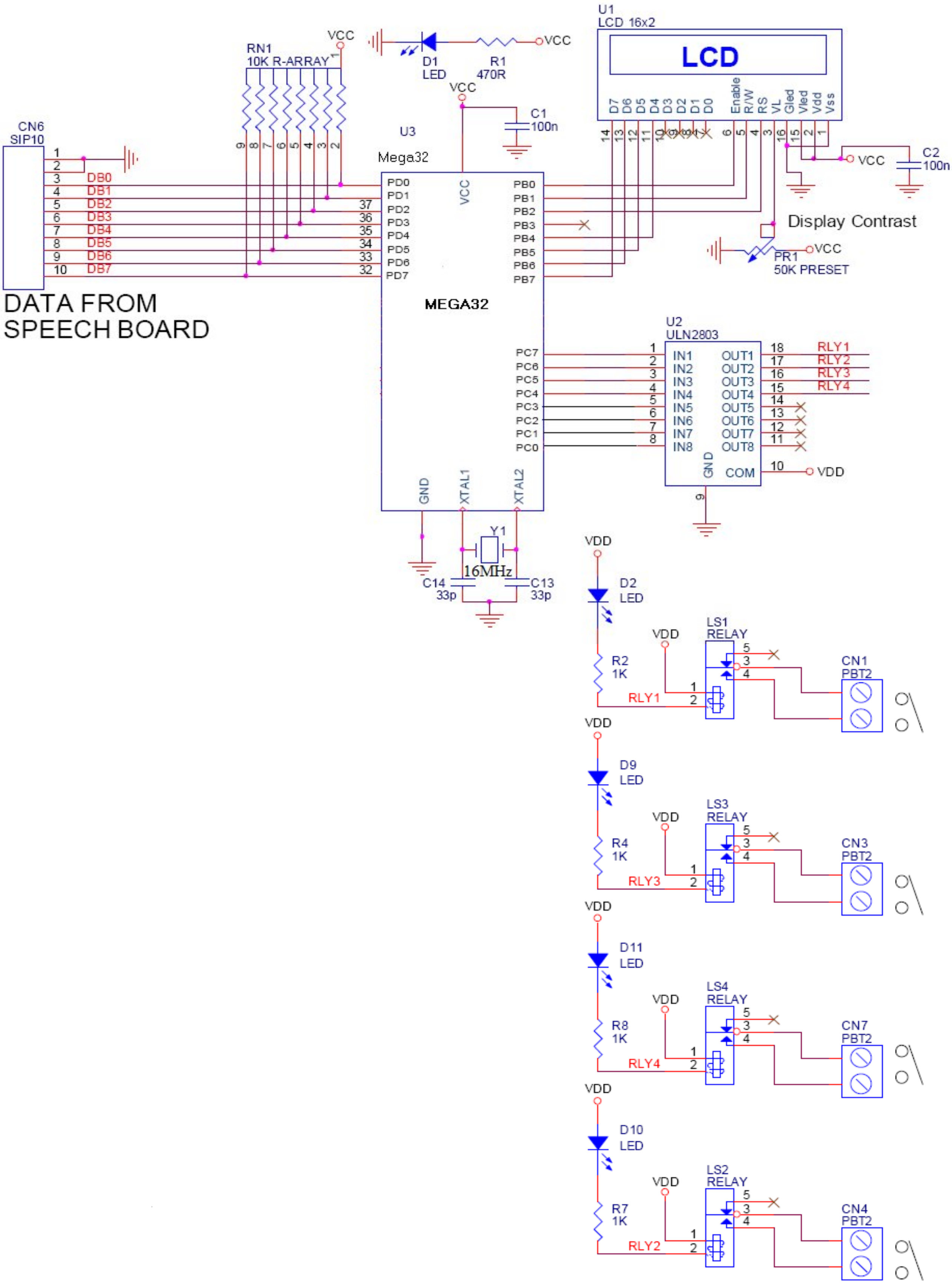
١. شريحة تمييز الصوت عالية التكامل من نوع COM.
٢. عملية التمييز فيها تعتمد على الشخص ،والكلمات معزولة عن بعضها.
٣. ذاكرة RAM يمكن أن توصل مباشرة مع الشريحة.
٤. عدد الكلمات الأعظمي التي يمكن تمييزها 40 كلمة.
٥. أطول فترة زمنية للكلمة 1.92 Sec.
٦. إمكانية توصيل عدة شرائح مع بعضها البعض.
٧. الميكروفون يمكن يوصل بشكل مباشر.
٨. زمن الاستجابة أقل 300 mSec.
٩. جهد التغذية 5V.

٤-٣ استخدام الدارة

يتم الضغط على المفتاح رقم 1 ثم المفتاح Train يتم النطق بالكلمة المراد تخزينها، تتحول الإشارة الصوتية بواسطة الميكروفون إلى إشارة جهد تشابهية؛ ثم تتحول الإشارة التشابهية إلى إشارة رقمية بواسطة محول ADC في الشريحة HM2007 ،من ثم تأتي المعالجة الرقمية للإشارة ، تحديد بداية ونهاية الكلمة واستخلاص المعلومات الضرورية المميزة للكلمة الصوتية، ويتم التخزين ضمن ذاكرة الـ RAM ، يتم تكرار هذه العملية حتى يتم تخزين عشرين كلمة.عندما يتم النطق بإحدى هذه الكلمات تقوم الشريحة بتنفيذ خوارزمية التمييز، و يتم إظهار رقم الكلمة على الشاشة ويتم استخدام هذه المعطيات للتحكم بإحدى أجهزة الخرج وهكذا.

في حال تم النطق بكلمة ما غير مخزنة فإنه يتم إظهار الرقم 77 يدل على الخطأ. وإذا كانت الكلمة طويلة جدا يتم إظهار الرقم 55.أما في حالة الكلمة قصيرة جدا يتم إظهار الرقم 66.

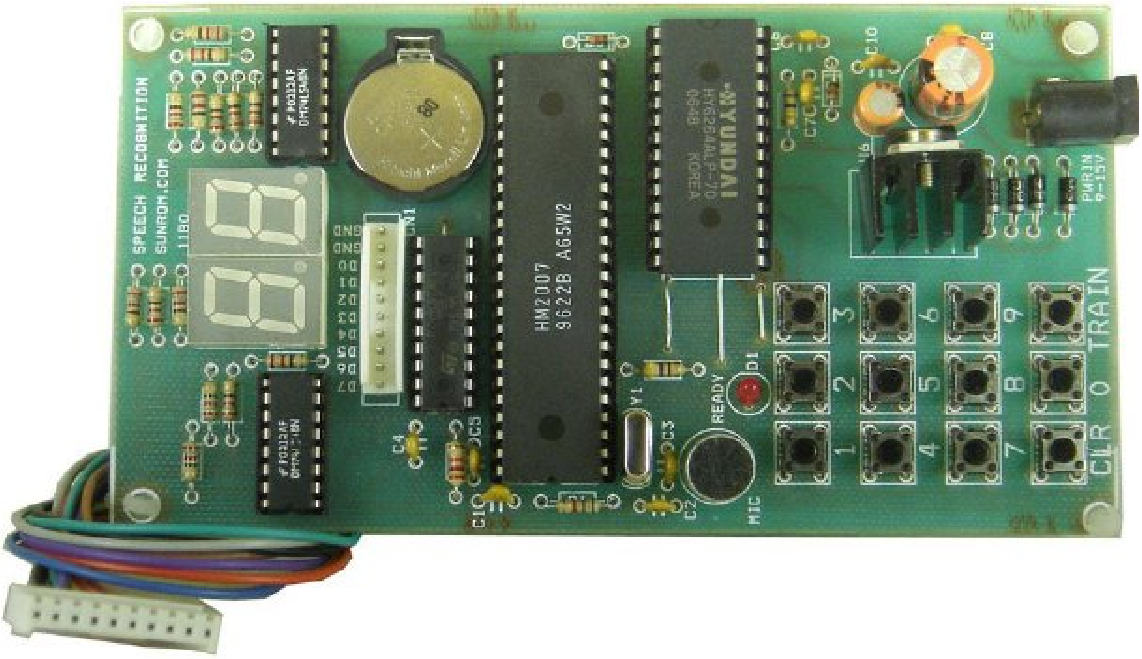
الشكل (٤-٣) يوضح قسم الميكروكونترولر من الدارة العملية،حيث يتم أخذ المعطيات من دارة التمييز واستخدامها في عملية التحكم ، تحتوي هذه الدارة على الميكروكونترولر MEG32 ،دارة متكاملة ULN2003 لقيادة الأحمال (كونتاكتور) شاشة إظهار LCD لإظهار ترتيب الكلمة، أربع كونتاكتورات يمكن استخدامها لتوصيل التغذية إلى محرك أو حمل آخر. يتم كتابة برنامج الميكروكونترولر بلغة الـ C الذي يمثل برنامج التحكم للدارة ،وهو مبين في الملحق A.



الشكل (٤-٣) دارة التحكم

٤–٤ اختبار دارة تمييز الصوت

الشكل (٤–٤) يوضح صورة لدارة تمييز الصوت. بإجراء عملية اختبار على دارة تمييز الصوت، تم اختبار سبع عشرة كلمة ، باستخدام صوت شخصين مختلفين وتم تخزين الكلمة لمرة واحدة وضمن غرفة مع قيمة الضجيج صغيرة 40dB-، والنتائج مبينة في الجدول (٤–٢).نلاحظ من هذه النتائج أن الدارة قادرة على تمييز كلمات من لغات مختلفة، وبنسبة تمييز كلية 86.64% للشخص الأول و88.17% للشخص الثاني؛ المتوسط الحسابي لنسبة التمييز 87.4% ، وتعتبر هذه النسبة جيدة.

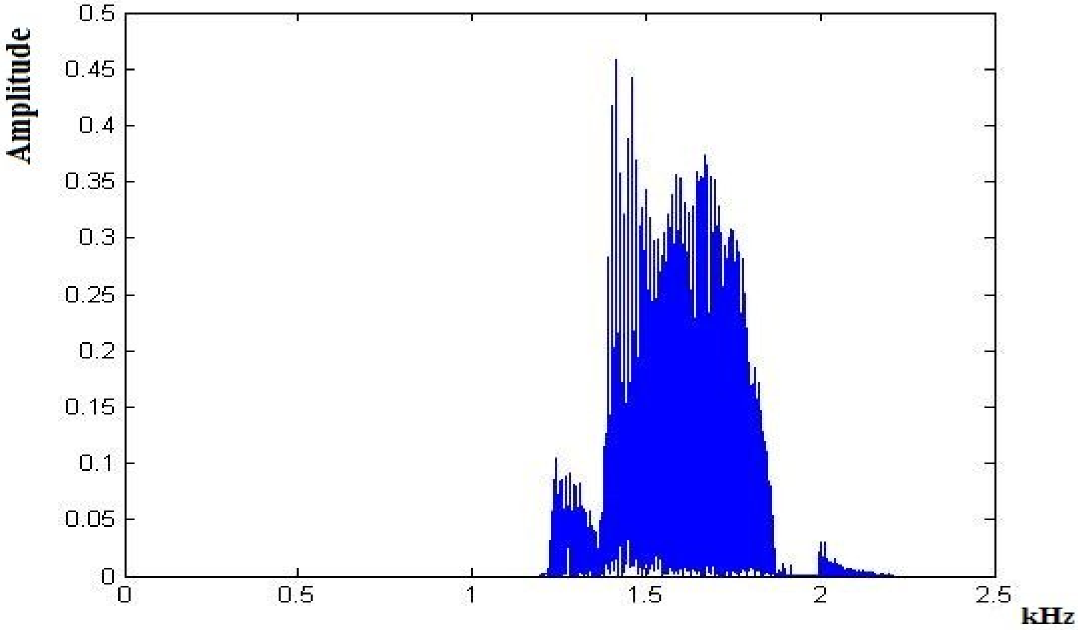


الشكل (٤–٤) يوضح صورة لدارة تمييز الصوت

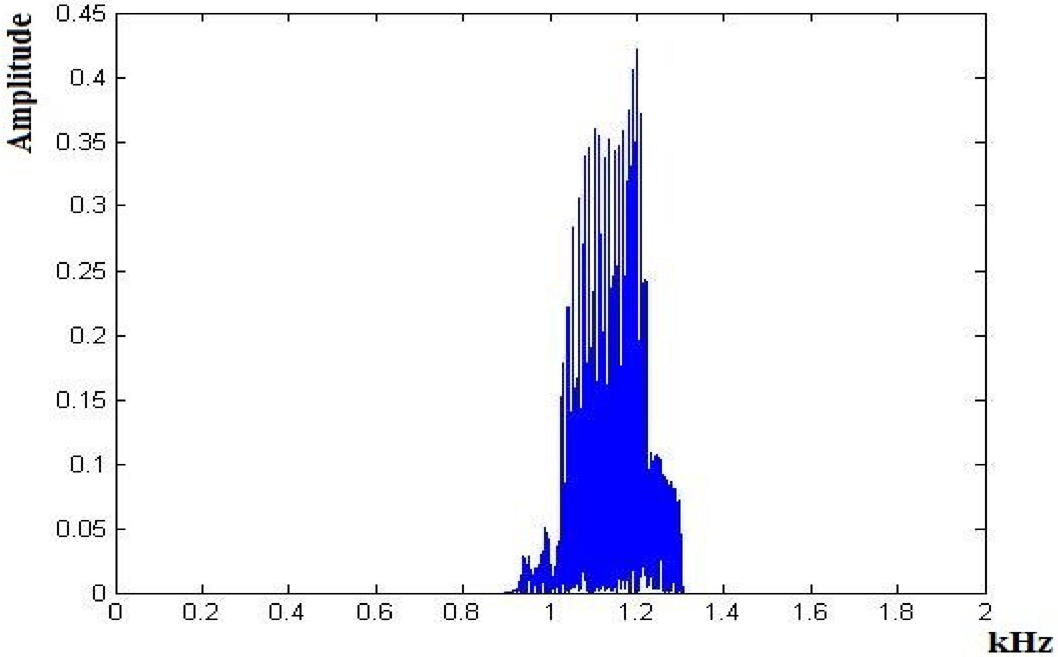
الجدول(٤-٢) ضمن غرفة من دون ضجيج بقيمة ضجيج 40dB-

الكلمة	عدد مرات التجريب	نسبة التمييز للشخص الأول	نسبة التمييز للشخص الثاني
Hello	20	65%	90%
Left	20	95%	85%
Right	20	90%	70%
Orange	20	65%	78%
Me	20	85%	95%
Car	20	95%	95%
Bare	20	85%	80%
Salad	20	80%	85%
Gat	20	98%	99%
Bid	20	99%	99%
أمام	20	75%	80%
يمين	20	85%	85%
يسار	20	80%	80%
ثانية	20	95%	100%
سبب	20	85%	80%
قال	20	98%	99%
جال	20	98%	99%

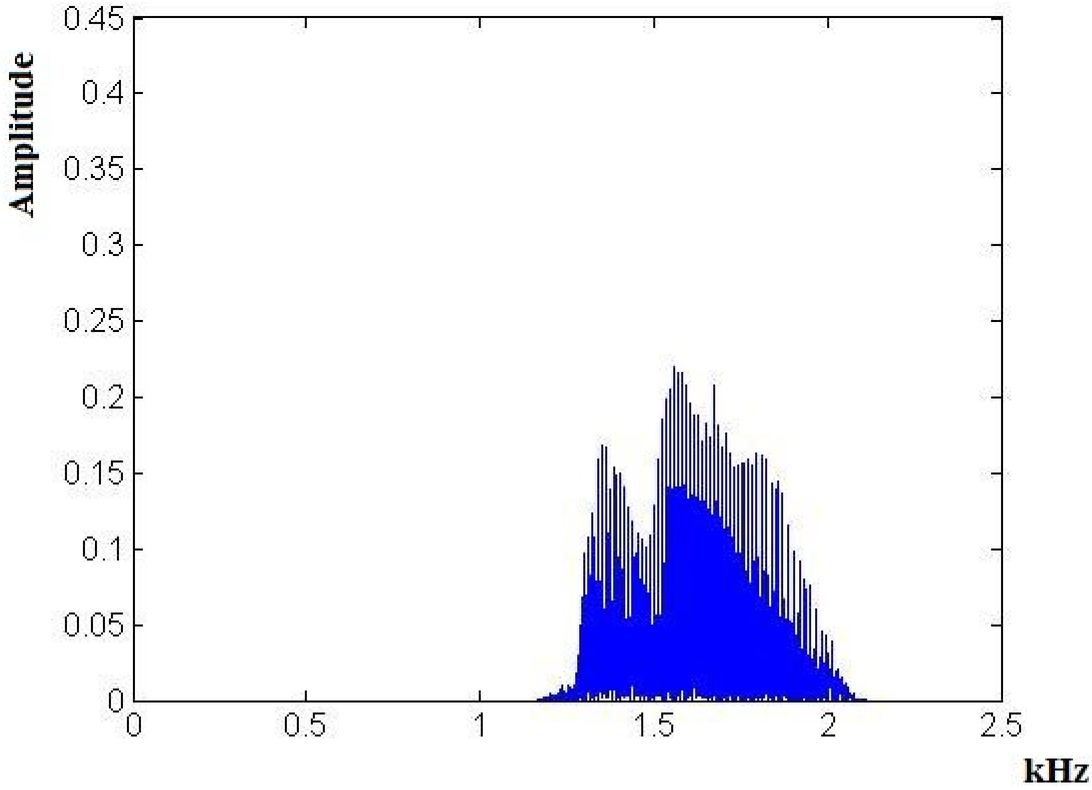
نلاحظ من هذه النتائج إن نسبة تمييز الصوت مختلفة وهي اكبر في الكلمات التي تحتوي على الحروف الانفجارية ولفهم ذلك سوف نوجد الطيف الترددي لبعض هذه الكلمات؛ في الشكل (٤-٤) يبين الطيف الترددي لكلمة (Gat) والشكل (٤-٥) لكلمة (Bid) نلاحظ بأن هذه الطيف الترددي لهذه الكلمات يحتوي على مطال واضح بالمقارنة مع الشكل (٤-٦) الذي يمثل الطيف الترددي لكلمة (Hello)حيث نلاحظ مطال الطيف الترددي فيها لا يتجاوز 0.25.



الشكل (٤-٤) يوضح الطيف الترددي لكلمة Gat



الشكل (٤-٥) يوضح الطيف الترددي لكلمة Bid



الشكل (٤-٦) يبين الطيف الترددي لكلمة Hello

لبيان تأثير الضجيج على دارة تمييز الصوت قمنا بتوليد ضجيج ابيض بواسطة برنامج Audio Editor Gold(Version 9.2.14) بقيمة 20dB- ؛كانت النتائج كما هو مبين في الجدول (٤-٣) إن نسبة تمييز الصوت الكلية أصبحت 87.26% نلاحظ إن هذه القيمة أصغر من قيمة نسبة تمييز الصوت السابقة ولكن بقليل وهذا يبين إن تأثير الضجيج على عمل الدارة قليل.

الجدول(٤-٣) ضمن غرفة مع ضجيج بقيمة 20dB-

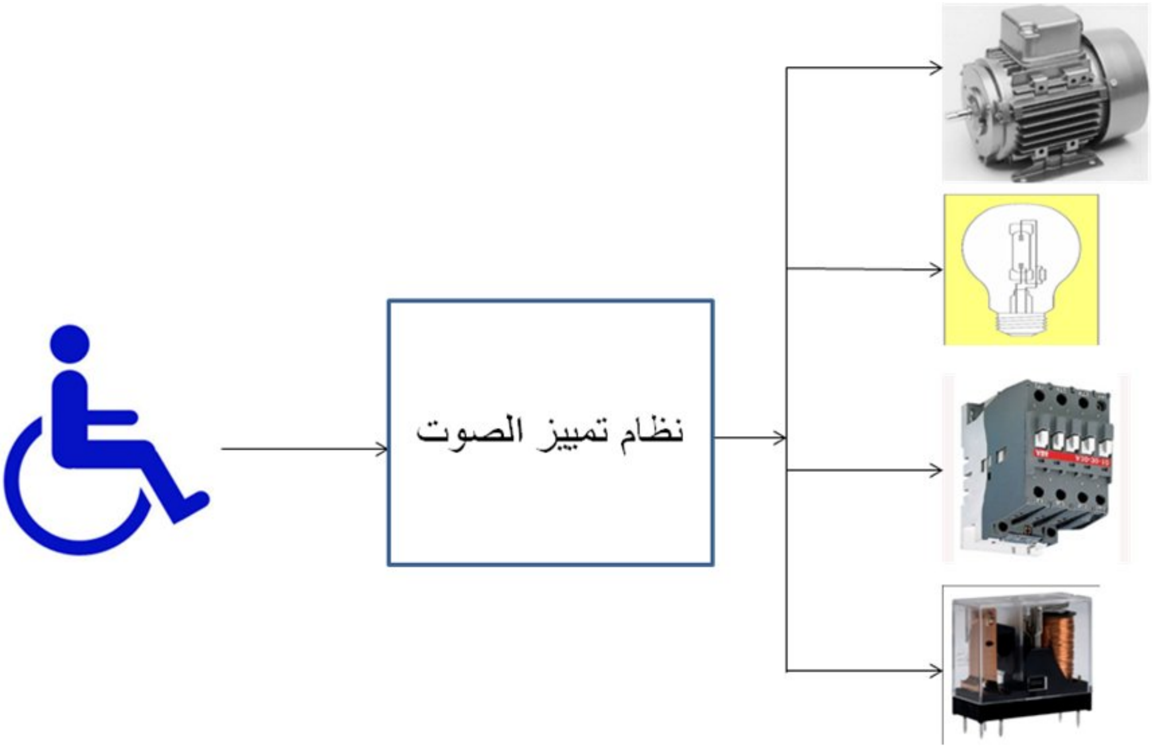
الكلمة	عدد مرات التجريب	نسبة التمييز للشخص الأول	نسبة التمييز للشخص الثاني
Hello	20	64%	90%
Left	20	95%	85%
Right	20	90%	68%
Orange	20	64%	77%
Me	20	85%	95%
Car	20	95%	95%
Bare	20	85%	80%
Salad	20	80%	85%
Gat	20	98%	99%
Bid	20	99%	99%
أمام	20	75%	80%
يمين	20	85%	85%
يسار	20	80%	80%
ثانية	20	95%	100%
سبب	20	85%	80%
قال	20	98%	99%
جال	20	98%	99%

٤-٥مقارنة النتائج

إن نسبة تمييز الصوت التي توصلنا إليها 87.4% جيدة ، حيث تم التوصل إلى نسبة تمييز الصوت 98.3% في بعض الأبحاث في جامعة كاجوا اليابانية [13]. ولكن يمكن أن نستخدم الكلمات التي تحتوي على الحروف الانفجارية للحصول على نسبة تمييز 99% وبالتالي يمكن الوصول إلى وثوقية عالية لهذه الدارة.

٤-٦ استخدام الدارة لمساعدة المعاقين

إن الدارة السابقة التي توصلنا إليها من خلال هذه الدراسة يمكن اعتبار مدخلها الصوت أي إن عملية التحكم بشكل كامل يتم عن طريق الصوت وخرجها يمكن أن يكون أما محرك أو LED ...الخ.والشكل (٤-٧)يبين ذلك.



الشكل (٤-٧) يوضح كيف يمكن مساعدة المعاق بطريقة التمييز الصوتي

الفصل الخامس
الخلاصة العامة والآفاق المستقبلية

٥-١ خلاصة العامة

إ ن هذا البحث يندرج تحت عنوان " إصدار الأوامر الصوتية من شخص معاق للتحكم بالأجهزة المحيطة " لذا قمنا بإيجاد خوارزمية تعتمد بشكل عام على المميزات الإحصائية للإشارة الصوتية .

حيث قدمنا من خلال الفصل الأول دراسة مقتضبة عن مفهوم الإشارة الصوتية وكيفية توليد الكلام عند الإنسان وذلك بعرض المخططات الأساسية لتوليد الكلام وبالإضافة لإلية انتشار الصوت في جهاز النطق عند الإنسان.

ثم تطرقنا في الفصل الثاني إلى دراسة مبادئ تحليل ومعالجة الإشارة لعملية تمييز الصوت" حيث تم في هذا الفصل عرض دراسة مفصلة لكافة المبادئ المستخدمة لتحليل الإشارة وكما تم عرض الخوارزميات التي يمكن استخدامها لاستخلاص المعلومات المفيدة في الإشارة الصوتية.

أما في الفصل الثالث فقد قمنا بدراسة خوارزميات تمييز الصوتي وتصميم الخوارزمية المناسبة لذلك.

وفي الفصل الرابع تم إجراء محاكاة للخوارزمية وتم عرض النتائج ومناقشتها .

وأخيرا تم دراسة الدارة العملية وتنفيذها للتحكم بأجهزة محيطة بالشخص المعاق .

وفي النهاية يمكننا تلخص نتائج هذا البحث بالنقاط الرئيسة التالية:

- دراسة الإشارة الصوتية من ناحية توليدها وانتشارها.
- دراسة الإشارة الرقمية وأساليب معالجتها.
- دراسة خوارزميات تمييز الصوت.
- إيجاد خوارزمية تعتمد على الخواص الإحصائية للإشارة الصوتية.
- إجراء دراسة للخوارزمية باستخدام بيئة MATLAB/SIMULINK .
- تنفيذ الدارة العملية باستخدام الشريحة HM2007 و ATMega32 .

٥-٢ الآفاق المستقبلية

إن الخوارزمية التي تمت دراستها تستخدم لتمييز الكلمات المفردة، تم الاعتماد على الخواص الإحصائية للإشارة (المتوسط الحسابي، الانحراف المعياري،عملية الترشيح...الخ)؛ لذلك فهي لا تحتاج إلى سرعة معالجة عالية نسبياً، ويمكن تطويرها بحيث تمييز كلمات متلاحقة وزيادة نسبة تمييز الصوت وكذلك زيادة عدد الكلمات التي يمكن تمييزها، مع إمكانية تنفيذها بشكل عملي،لكن يجب استخدام سرعة معالجة عالية.

٥-٣ المنشورات

تم نشر بحث بعنوان:" إصدار الأوامر الصوتية من شخص معاق للتحكم بالأجهزة المحيطة "
في مجلة بحوث جامعة حلب العدد ٨٢ لعام ٢٠١٠.

الملحقات

الملحق A

يمثل برنامج دارة التحكم

```
#include <mega32.h>
#include <math.h>
#include <stdio.h>
#include <stdlib.h>
#include <delay.h>
#include <lcd.h>
#asm
.equ __lcd_port=0x18 ;PORTB
#endasm
unsigned char lastdata, datanow;
unsigned char*str;
void main(void) {
PORTD=0x00;
DDRD=0xff;
PORTA=0xff;
DDRA=0x00;
// -=-=- Intialize variables -=-=-=
PORTD.0 = 0;
PORTD.1 = 0;
PORTD.2 = 0;
PORTD.3 = 0;
PORTD.4 = 0;
PORTD.5 = 0;
PORTD.6 = 0;
PORTD.7 = 0;
// -=-=- Intialise -=-=-=
lcd_init(16);
// -=-=- Welcome LCD Message -=-=-=
lcd_clear();
lcd_gotoxy(0,0); // 1st Line of LCD
lcd_putsf("Speech Test");
lcd_gotoxy(0,1); // 2nd Line of LCD
lcd_putsf("System");
delay_ms(5000); // 5 sec
lcd_clear();
lcd_gotoxy(0,0); // 1st Line of LCD
lcd_putsf("Train: 1-4 key >");
lcd_gotoxy(0,1); // 2nd Line of LCD
```

```
lcd_putsf("Train>Speak Now");
delay_ms(500);
// -=-=- Program Loop -=-=-=
lastdata=0xff;
lcd_init(16);
while(1)
{delay_ms(1);

datanow=PINA; // read data from speech board
//itoa(datanow,str);

//lcd_clear();
//lcd_gotoxy(0,0) ;
//lcd_puts(str);
//delay_ms(300);

if(lastdata!=datanow) // if there is new data then,
{
lastdata=datanow;
switch(lastdata)
{
case 0x55:
lcd_clear();
lcd_gotoxy(0,0); // 1st Line of LCD
lcd_putsf("Speech too Long");
lcd_gotoxy(0,1); // 2nd Line of LCD
lcd_putsf("Try Again!");
break;
case 0x66:
lcd_clear();
lcd_gotoxy(0,0); // 1st Line of LCD
lcd_putsf("Speech too Short");
lcd_gotoxy(0,1); // 2nd Line of LCD
lcd_putsf("Try Again!");
break;
case 0x77:
lcd_clear();
lcd_gotoxy(0,0); // 1st Line of LCD
lcd_putsf("No Match");
lcd_gotoxy(0,1); // 2nd Line of LCD
lcd_putsf("Try Again!");
break;
case 0x01:
if(PORTD.7==1)
```

```
PORTD = 0;
else
PORTD = 128;
lcd_clear();
lcd_gotoxy(0,0); // 1st Line of LCD
lcd_putsf("SPEECH: ONE");
break;
case 0x02:
if(PORTD.6==1)
PORTD = 0;
else
PORTD =64;
lcd_clear();
lcd_gotoxy(0,0); // 1st Line of LCD
lcd_putsf("SPEECH: TWO");
break;
case 0x03:
if(PORTD.5==1)
PORTD = 0;
else
PORTD = 32;
lcd_clear();
lcd_gotoxy(0,0); // 1st Line of LCD
lcd_putsf("SPEECH: THREE");
break;
case 0x04:
if(PORTD.4==1)
PORTD = 0;
else
PORTD = 16;
lcd_clear();
lcd_gotoxy(0,0); // 1st Line of LCD
lcd_putsf("SPEECH: FOUR");
break;
case 0x05:
if(PORTD.3==1)
PORTD = 0;
else
PORTD = 8;
lcd_clear();
lcd_gotoxy(0,0); // 1st Line of LCD
lcd_putsf("SPEECH: FIVE");
break;
case 0x06:
if(PORTD.2==1)
```

```
PORTD = 0;
else
PORTD =4;
lcd_clear();
lcd_gotoxy(0,0); // 1st Line of LCD
lcd_putsf("SPEECH: SIX");
break;
case 0x07:
if(PORTD.1==1)
PORTD = 0;
else
PORTD = 2;
lcd_clear();
lcd_gotoxy(0,0); // 1st Line of LCD
lcd_putsf("SPEECH: SEVEN");
break;
case 0x08:
if(PORTD.4==1)
PORTD = 0;
else
PORTD = 1;
lcd_clear();
lcd_gotoxy(0,0); // 1st Line of LCD
lcd_putsf("SPEECH: EIGHT");

}
}
}
}
```

الملحق B

يبين معلومات عن الشريحة HM2007

HM2007
SPEECH RECOGNITION

GENERAL DESCRIPTION

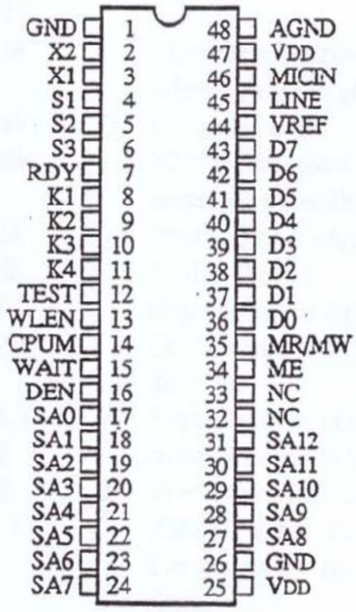
HM2007 is a single chip CMOS voice recognition LSI circuit with the on-chip analog front end, voice analysis, recognition process and system control functions. A 40 isolated-word voice recognition system can be composed of external microphone, keyboard, 64K SRAM and some other components. Combined with the microprocessor, an intellengent recognition system can be built.

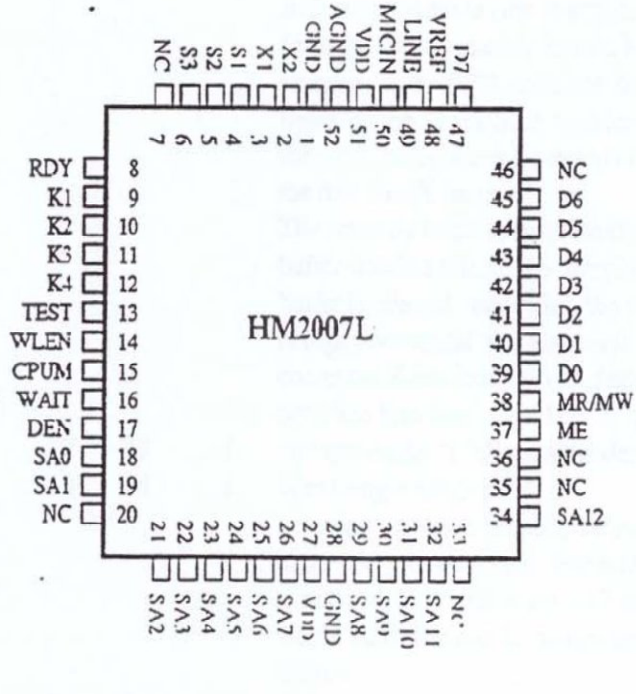
FEATURES

- Single chip voice recognition CMOS LSI.
- Speaker-dependent isolates-word recognition system.
- External 64K SRAM can be connected directly.
- Maximum 40 words can be recognized for one chip.
- Maximum 1.92 sec of word can be recognized.
- Multiple-chip configuration is possible.
- A microphone can be connected directly.
- Two control mode is supported: Manual mode and CPU mode.
- Response time : less than 300 ms.
- 5V single power supply.
- 48-pin PDIP, 52 pin PLCC, 48 pad bare chip.

PIN CONFIGURATIONS

HM2007P





PIN DESCRIPTIONS

Symbol	Pin No. PDIP 48L	Pin No. PLCC 52L	I/O	Function
Vref	44	48	I	The voltage reference input of internal ADC. Supply the reference voltage of the internal A/D converter.
LINE	45	49	O	For testing only.
MICIN	46	50	I	Microphone connect pin. A microphone should be connected via a coupling capacitor and resistor.
V_{DD}	47	51		Positive power supply.
AGND	48	52		Analog Ground.
GND	1	1		Negative power supply.
X2,X1	2,3	2,3	I	Crystal connect pin. A 3.58 MHz crystal is connected to these pin.
S1,S2 S3	4,5 6	4,5 6	I/O	Keypad scaning pin for manual mode and the read/write control Pins in the CPU mode.
RDY	7	8	O	Voice input ready indicator. Active low output.When HM2007 is ready for the voice input in training or recognition mode,a low signal is sent.If the chip is busy, a high signal is sent.
K1,K2 K3,K4	8-11	9-12	I/O	The keypad input pin in the manual mode and the bidirectional data bus (K-bus)in the CPU mode. In the manual mode , the four pins combined with S1 to S3 form the keypad scaning circuit. Maximum12 keys can be scanned. In the CPU mode, the data bus direction is determined by the S2 and S3.A high level signal that appears in the pin S2 will place the content of internal register onto to the data bus.(K-bus). The data may be come from the status register or the output buffer which is selected by the pin S1. If S1 is high, output buffer is selected, otherwise, the status register is selected. A high level signal that appears in the pin S3 will place the content of K-bus into the input register. Note that user can not place high level signal on S2 and S3 simultaneously.
TEST	12	13	I	"H":test mode. "L":Normal mode.
WLEN	13	14	I	Word length select pin. Selecting the voice length to be recognized. When set to high, 1.92 sec is selected. Internally pull low for 0.9 sec is selected. Note that when 1.92 sec is selected,only 20 words maximum can be recognized if 8K-byte memory is used.
CPUM	14	15	I	CPU mode select pin. Internally pull low for manual mode. When set to high,CPU mode is selected.

Symbol	Pin No.		I/O	Function
	PDIP 48L	PLCC 52L		
WAIT	15	16	I	Waiting control input. Active low input. When this pin is set to "L" and manual mode is selected, HM2007 will enter the waiting state and do not accept voice input until this pin back to "H". For CPU mode, when HM2007 is ready to get voice input, if this pin is set to "L", HM2007 will skip the voice input process and enter the get-command process.
DEN	16	17	O	Data enable signal. When the recognition or training process is complete, the chip will place its response on the data bus D0 to D7 and which can be latched onto external devices by this pin.
SA0,SA1 SA2-SA7 SA8-SA11 SA12	17-24 27-31	18,19 21-26 29-32 34	O	External memory address bus. The bus is used as an external memory address when ME pin is active.
V_{DD}	25	27		Positive power supply.
GND	26	28		Negative power supply.
NC	32,33	35,36 7,20 33,46		
ME	34	37	O	Memory enable pin. Active low output. This pin will send the memory enable signal to the external SRAM. This pin can be connected directly to the CE pin of 6264 SRAM.
MR/MW	35	38	O	Memory read/write select pin. Read/write control signal of the external SRAM. This pin can be connected directly to the R/W pin of 6264 SRAM.
D0-D6 D7	36-42 43	39-45 47	I/O	External memory data bus(D-bus). The bus is used as an external memory I/O bus when ME pin is active and used as output response bus when DEN pin is active.

FUNCTION DESCRIPTIONS

There are two operation mode which are provided by HM2007.

A). Manual mode.

In this operation mode, a keypad, a SRAM and other components may be connected to HM2007 to build a simple recognition system (See application circuit). The type of SRAM can be used is a 8K-byte memory.

a). Power on.

When the power is on.HM2007 will start its initialization process. If WAIT pin is "L", HM2007 will do the memory check to see whether the external 8K byte SRAM is perfect or not.
If WAIT pin is "H", HM2007 will skip the memory check process.
After the initial process is done, HM2007 will then move into recogition mode.

b). Recognition Mode.

i). WAIT pin "H"

In this mode, the RDY is set to low and HM2007 is ready to accept the voice input to be recognized.
When the voice input is detected, the RDY will return to high and HM2007 begins its recognition process.It is recommended that user train the word pattern before the beginning of the recognition operation, otherwise the result will be unpredictable. After the recognition process is completely, the result will appear on the D-bus with the pin DEN active.
Table 1 shows the list of the output content.
The data on the data bus is a decimal code in binary format.

D7	D6	D5	D4	D3	D2	D1	D0	Description
0	0	0	0	0	0	0	0	Power on
		A				B		Word AB
0	1	0	1	0	1	0	1	Voice too long
0	1	1	0	0	1	1	0	Voice too short
0	1	1	1	0	1	1	1	Not Match

Table 1 : Content of the D-bus output.

Note 1 : A is the binary code in the range 0 to 4, and B is the binary code in the range 0 to 9.
Note 2 : If WLEN is high, the maximum word length is 1.92 sec.

ii).WAIT pin "L"

In this mode, no voice input is accepted until WAIT pin back to "H" state.

c). Training or clearing one pattern

Two operation are included during this time, 1) clearing trained pattern and 2).training new patern.
To clear or train the voice pattern, one must select the word number to process first. The number of word is composed of two digits. The two digits are entered into HM2007 through keypad one digit a time. If more than two digits are entered, only the last two digits are valid. When number key is pressed, the number of key will be echoed to the D-bus.

When the word number is entered, press the function key to choose the operation function. If function key CLR is pressed. the corresponding word pattern will be cleared and then HM2007 will return its recognition mode. If the function key TRN is pressed, HM2007 will begin its training process. At the beginning of training process, if WAIT pin is "H", HM2007 will send a low level signal to RDY to indicate that HM2007 is ready to accept voice input. If WAIT pin is "L", no voice input will be detected until WAIT pin back to "H". After available voice input to HM2007, HM2007 will return to its recognition mode and send a low level signal to RDY to indicate that HM2007 is ready for voice input to do the recognition process.

For example.
2 4 TRN → training the 24th pattern.
0 1 CLR → clearing the first pattern.
1 3 2 6 TRN → training the 26th pattern.

d). Clear all pattern.

If the number key 99 is entered and the CLR is pressed, all the patterns in the memory will be cleared by HM2007.

B). CPU control mode.

The CPU mode provides several functions: RECOG, TRAIN, RESULT, UPLOAD, DOWNLOAD, RESET and which will be described later. In this mode, the K-bus is used as a bidirectional data bus between the external controller and HM2007 and S1 to S3 as the R/W control pins.
Table 2 is the summary of the CPU command.The command contains two parts. 1). the command code and 2). the number of the word to be processed if needed.

command	code	word #(L)	word #(H)
RECOG	0001		
TRAIN	0010	B3 B2 B1 B0	0 0 B5 B4
RESULT	0100		
UPLOAD	0101	B3 B2 B1 B0	0 0 B5 B4
DOWNLOAD	0110	B3 B2 B1 B0	0 0 B5 B4
RESET	0111		

Table 2. Command for CPU mode.

There are three registers in HM2007, one input buffer register, one status register and one output buffer register. The first is a write-only register and the last two are read-only registers. If S1 pin is high, the data read from the K-bus will come from the output buffer register. If S1 pin is low, the data read from the K-bus will come from the status register. S2 and S3 are R/W control signals. If S2 is high. it's in a read cycle and the external controller can read data form the K-bus. If S3 is high, it's in a write cycle and external controller can write data into the input buffer. Note that S2 and S3 can not be high simultaneously and the state of S1 will be ignored during a write cycle.
The status register as shown. reflect the current status of HM2007 for the CPU control mode.

المراجع

[1] Holmes W.,2001"**Speech synthesis and Recognition**" .CRC Press Second edition ,London ,350 pages.

[2] Fry D., 1959" **Theoretical Aspects Of Mechanical Speech ecognition**". CRC Press First edition ,London,1489 pages.

[3] Rabiner L .,1993"**Fundamentals of speech recognition**". Prentice-Hall First edition .USA.507Pages

[4] Steven W. , 1999"**The Scientist and Engineer's Guide to Digital Signal Processing**" . California Technical Publishing, Second Edition ,USA,400pages

[5] Rabiner L.,1975"**Algorithm for determining the endpoints of isolated utterances The Bell System Technical**". Journal , Vol 54.

[6] Dattorro J.,2005" **Convex Optimization & Euclidean Distance Geometry**",MEBOO publishing First edition USA,485pages

[7] Sarma V., 1978" **Studies on pattern recognition approach to voiced-unvoiced-silence classification**". Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '78. , Volume: 3, Pages: 1 - 4.

[8] Flanagan L., 1972"**Speech Analysis, Synthesis, and Perception**". Second edition., Springer-Verlag, New York,350 pages.

[9] Picone J., 1993"**FUNDAMETALS OF SPEECH RECOGNITION:A SHORT COURSE**".Institute for Signal and Information Processing Department of Electrical and Computer Engineering Mississippi State University, Second edition,USA,150 pages.

[10] Martin A.,2001 "**Robust speech / non-speech detection using LDA applied to MFCC**", in Proc. ICASSP2001}, vol. 1, 2009, pp. 237–240.

[11] Richard O., 2009 "**Pattern Classification**", A Wiley-interscience publication, Second Edition,London,370 pages.

[12] ISECS International Colloquium on Computing, Communication, Control, and Management IEEE CCCM 2009,"**Design and Realization of an Intelligent Access Control System Based on Voice Recognition**", China.
[13]SICE Annual Conference 2007,Sept. 17-20, 2007 .,"**Voice Controlled Intelligent Wheelchair**" Kagawa University, Japan.

Summary

Voice Recognition is a computerized process. It receives speech recording as its input, and produces a command as its output. In the computing world, Voice Recognition is an unsolved problem; nobody has yet demonstrated the ability to do it as well as humans can. This thesis looks at the Voice Recognition process, demonstrating some ways to help handicapped to control the environment. The signal was obtained from a microphone and transformed into a digital signal, then filtered using ten digital filters in order to extract the beneficial information that represent the fingerprint of a voice word to be stored in memory. This operation called training phase. When we talk, the system generates template fingerprint and compares them with stored fingerprints to get the result and produce the required command output .The system is simulated using MATLAB/SIMULINK environment and the results were acceptable.

Testimony

We witness that the described work in this thesis is the result of scientific search conducted by the candidate Eng. Ramadan Khalil Almahmoud under supervision of Dr. Yasser Estanbouli Assistant Professor at the Faculty of Electrical & Electronic Engineering, University of Aleppo, and Dr.Omar Sardar (assistant supervisor) Assistant Professor at the Faculty of Electrical & Electronic Engineering University of Aleppo. Any other references mentioned in this work are documented in the text of the thesis.

Candidate	Assistant supervisor	Main supervisor
Eng.Ramadan Almahmoud	Dr. Omar Sardar	Dr.Yasser Estanbouli

Declaration

I hereby certify that this work has not been accepted for any degree or it is not submitted to any other degree.

Candidate

Eng. Ramadan Almahmoud

Aleppo University
Faculty of Electrical & Electronic Engineering
Electronic Engineering Department

Design of Handi Capped Sound Based System to Control The Environment

This Thesis is Submitted to Obtain the Master Degree in Electrical Engineering

By
Dipl. Eng. Ramadan Almahmoud
Master Student
Department of Electronic Engineering
Faculty of Electrical & Electronic Engineering
University of Aleppo

Supervised By

Dr. Yasser Estanbouli Assistant Professor at the Faculty of Electrical & Electronic Engineering University of Aleppo	Dr. Omar Sardar Assistant Professor at the Faculty Electrical of Electrical & Electronic Engineering University of Aleppo

Aleppo University
Faculty of Electrical & Electronic Engineering
Electronic Engineering Department

Design of Handi Capped Sound Based System to Control The Environment

This Thesis is submitted to obtain the Master Degree in Electronic Engineering

By

Eng. Ramadan Almahmoud
Postgraduate Student
Department of Electronic Engineering
Faculty of Electrical & Electronic Engineering
Aleppo University